B1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/031571 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/31357

(22) International Filing Date: 2 October 2002 (02.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/327,454	5 October 2001 (05.10.2001)	US
60/327,917	9 October 2001 (09.10.2001)	US
60/328,029	9 October 2001 (09.10.2001)	US
60/328,056	9 October 2001 (09.10.2001)	US
60/328,849	12 October 2001 (12.10.2001)	US
60/329,414	15 October 2001 (15.10.2001)	US
60/330,142	17 October 2001 (17.10.2001)	US
60/341,058	22 October 2001 (22.10.2001)	US
60/343,629	24 October 2001 (24.10.2001)	US
60/349,575	29 October 2001 (29.10.2001)	US
60/346,357	1 November 2001 (01.11.2001)	US
60/391,342	25 June 2002 (25.06.2002)	US
10/262,445	1 October 2002 (01.10.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/327,454 (CIP)
Filed on	5 October 2001 (05.10.2001)
US	60/327,917 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,029 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,056 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,849 (CIP)
Filed on	12 October 2001 (12.10.2001)
US	60/329,414 (CIP)
Filed on	15 October 2001 (15.10.2001)
US	60/330,142 (CIP)
Filed on	17 October 2001 (17.10.2001)
US	60/341,058 (CIP)
Filed on	22 October 2001 (22.10.2001)
US	60/343,629 (CIP)
Filed on	24 October 2001 (24.10.2001)
US	60/349,575 (CIP)
Filed on	29 October 2001 (29.10.2001)
US	60/346,357 (CIP)
Filed on	1 November 2001 (01.11.2001)
US	60/391,342 (CIP)
Filed on	25 June 2002 (25.06.2002)
US	Not furnished (CIP)
Filed on	1 October 2002 (01.10.2002)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ALSOBROOK, John, P., II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **CATTERTON, Elina** [FI/US]; 584 Boston Post Road, Madison, CT 06443 (US). **CHANT, John, S.** [CA/US]; 76 Peddlers Lane, Branford, CT 06405 (US). **CHAUDHURI, Amitabha** [IN/US]; 99 Harbor Avenue, Madison, CT 06443 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **GIOT, Loic** [FR/US]; 99 Country Way, Madison, CT 06443 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **MEZES, Peter, S.** [CA/US]; 7 Clark's Lane, Old Lyme, CT 06371 (US). **MILLET, Isabelle** [FR/US]; 74 Carrington Avenue, Milford, CT 06460 (US). **OOI, Chean, Eng** [US/US]; 14 Flax Mill Hollow, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **RIEGER, Daniel, K.** [DE/US]; 10A McKinnel Court, Branford, CT 06405 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, Number 1, New Haven, CT 06511 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Haihong** [CN/US]; 2269 Long Hill Road, Guilford, CT 06437 (US). **ZHONG, Mei** [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US).

[Continued on next page]

(54) Title: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies that immunospecifically bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the novel polypeptide, polynucleotide, or antibody specific to the polypeptide. Vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same are also included. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 03/031571 A2

(74) **Agents: ELRIFI, Ivor, R.** et al.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

FIELD OF THE INVENTION

The present invention relates to nucleic acids encoding proteins that are new members of the following protein families: intracellular protein-like Proteins, Sorting Nexin 6-like Proteins, 2310038H17RIK membrane (TmSP) protein-like Proteins, 573045I09RIK cyclin-like Proteins, cMob5 cancer specific proteins, LRP16 protein-like Proteins, Phosphatidylethanolamine-binding protein-like Proteins, Immunoglobulin-like LRR-domain containing Proteins, NUMB binding protein LNXp80-like Proteins, Zinc Finger Protein-like Proteins, Actin-Binding Protein Frabin-Alpha-like Proteins, Actin related protein 2/3 complex subunit 1A-like Proteins, Hepatocellular Carcinoma Autoantigen - like Proteins, Hematopoietic Stem/Progenitor Cells Protein MDS029 -like Proteins, TRAP-delta-like Proteins, INTSIG-5-like WD-40 repeats containing protein-like Proteins, Ferritin light chain-like Proteins, Leucine-rich protein 130-like Proteins, tumor protein p53-binding protein 2 - like Proteins.

Included in the invention are polynucleotides and the polypeptides encoded by such polynucleotides, as well as vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using the same. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The present invention is based in part on nucleic acids encoding proteins that are members of the following protein families: intracellular protein-like Proteins, Sorting Nexin 6-like Proteins, 2310038H17RIK membrane (TmSP) protein-like Proteins, 573045I09RIK cyclin-like Proteins, cMob5 cancer specific proteins, LRP16 protein-like Proteins, Phosphatidylethanolamine-binding protein-like Proteins, Immunoglobulin-like LRR-domain containing Proteins, NUMB binding protein LNXp80-like Proteins, Zinc Finger Protein-like Proteins, Actin-Binding Protein Frabin-Alpha-like Proteins, Actin related protein 2/3 complex subunit 1A-like Proteins, Hepatocellular Carcinoma Autoantigen - like Proteins, Hematopoietic Stem/Progenitor Cells Protein MDS029 -like Proteins, TRAP-delta-like Proteins, INTSIG-5-like WD-40 repeats containing protein-like Proteins, Ferritin light chain-like Proteins, Leucine-rich protein 130-like Proteins, tumor protein p53-binding protein 2 - like Proteins. The novel polynucleotides and polypeptides are referred to herein as NOV1a, NOV1b, NOV2a, NOV2b, NOV3a, NOV3b, NOV3c, NOV4a, NOV4b, NOV5a, NOV6a, NOV6b, NOV7a, NOV8a, NOV9a, NOV10a, NOV11a, NOV12a, NOV13a, NOV14a, NOV15a, NOV16a, NOV16b, NOV17a, NOV17b, NOV18a, NOV18b, NOV18c, NOV19a, NOV19b, NOV20a, NOV20b, NOV20c. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid disclosed in SEQ ID NO:2n-1, wherein n is an integer between 1 and 33. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or

derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2n, wherein n is an integer between 1 and 33. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NO:2n-1, wherein n is an integer between 1 and 33. Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide that includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NO:2n-1, wherein n is an integer between 1 and 33) or a complement of said oligonucleotide.

The invention also encompasses isolated NOVX polypeptides (SEQ ID NO:2n, wherein n is an integer between 1 and 33). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid

probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33, the method including providing a cell expressing the polypeptide of the invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, e.g., adrenoleukodystrophy, congenital adrenal hyperplasia, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, lupus erythematosus, psoriasis, lung disorders, liver disorders, rheumatoid arthritis, osteoarthritis, Crohn's disease, ulcerative colitis, inflammatory bowel disease, asthma, allergies, chronic obstructive pulmonary disease, immunodeficiencies, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, schizophrenia, depression, anxiety, pain, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, asthma, emphysema, scleroderma, adult respiratory distress syndrome (ARDS), lymphedema, graft versus host disease (GVHD), pancreatitis, obesity,

ulcers, anemia, ataxia-telangiectasia, cancer, trauma, viral infections, bacterial infections, parasitic infections; and conditions related to transplantation, neuroprotection, fertility, or regeneration (in vitro and in vivo) and/or other pathologies and disorders of the like. Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing conditions including, *e.g.*, those associated with homologs of a NOVX sequence, such as those listed in Table A.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal that recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test

animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identity the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

NOVX nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOVX substances for use in therapeutic or diagnostic methods. These NOVX antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOVX proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These NOVX proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

The NOVX nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides, antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
NOV1a	CG116579-01	1	2	anticancer protein, intracellular protein
NOV1b	CG116579-02	3	4	anticancer protein, intracellular protein
NOV2a	CG126119-01	5	6	Sorting nexin 6 (TRAF4-associated factor 2)
NOV2b	CG126119-02	7	8	Sorting nexin 6-like Protein
NOV3a	CG137623-01	9	10	Hepatocellular carcinoma-associated antigen HCA557b; 2310038H17RIK membrane (TmSP) protein
NOV3b	CG137623-02	11	12	antigen HCA557b
NOV3c	CG137623-03	13	14	antigen HCA557b
NOV4a	CG137687-01	15	16	573045I09RIK cyclin-like Protein
NOV4b	CG137687-02	17	18	cyclin-like Proteins
NOV5a	CG143198-01	19	20	1500011J06Rik protein; Nuclear protein
NOV6a	CG144756-01	21	22	cMob5 cancer specific protein
NOV6b	CG144756-02	23	24	Suppression of tumorigenicity 16 protein
NOV7a	CG145473-01	25	26	LRP16 protein
NOV8a	CG145988-01	27	28	Phosphatidylethanolamine-binding protein; Prostatic binding protein
NOV9a	CG146452-01	29	30	LRR domain-containing Protein; Immunoglobulin protein
NOV10a	CG146731-01	31	32	Membrane Binding-like Protein; NUMB binding protein LNXp80; multi-PDZ-domain-containing protein
NOV11a	CG147048-01	33	34	Zn Finger Protein; RING finger protein 18
NOV12a	CG147246-01	35	36	Actin-Binding Protein Frabin-Alpha; GEF
NOV13a	CG147651-01	37	38	Suppressor of profilin/p41 of Actin related protein 2/3 complex (subunit 1A)
NOV14a	CG149303-01	39	40	Hepatocellular Carcinoma Autoantigen
NOV15a	CG149312-01	41	42	hematopoietic stem/progenitor cell protein MDS029
NOV16a	CG150951-01	43	44	Translocon-associated protein, delta subunit precursor (TRAP-delta)
NOV16b	CG150951-02	45	46	TRAP-delta-like Proteins
NOV17a	CG173328-01	47	48	intracellular signaling (INTSIG-5) protein; WD-40 repeat containing protein
NOV17b	CG173328-02	49	50	WD-40 repeats containing protein
NOV18a	CG56101-01	51	52	ferritin light chain protein
NOV18b	CG56101-03	53	54	Ferritin light chain protein
NOV18c	CG56101-02	55	56	ferritin light chain protein
NOV19a	CG56620-01	57	58	leucine-rich protein (LRP 130) (GP130)
NOV19b	CG56620-02	59	60	Leucine-rich protein
NOV20a	CG59323-01	61	62	Tumor suppressor p53-binding protein 2 (TP53BP2) (Bcl2-binding protein); APS-2 apoptosis stimulating protein 2 (ASSP)
NOV20b	CG59323-03	63	64	TP53BP2; ASSP
NOV20c	CG59323-02	65	66	TP53BP2; ASSP

Table A indicates the homology of NOVX polypeptides to known protein families. Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to: e.g., adrenoleukodystrophy, congenital adrenal hyperplasia, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, lupus erythematosus, psoriasis, lung disorders, liver disorders, rheumatoid arthritis, osteoarthritis, Crohn's disease, ulcerative colitis, inflammatory bowel disease, asthma, allergies, chronic obstructive pulmonary disease, immunodeficiencies, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, schizophrenia, depression, anxiety, pain, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, asthma, emphysema, scleroderma, adult respiratory distress syndrome (ARDS), lymphedema, graft versus host disease (GVHD), pancreatitis, obesity, ulcers, anemia, ataxia-telangiectasia, cancer, trauma, viral infections, bacterial infections, parasitic infections; and conditions related to transplantation, neuroprotection, fertility, or regeneration (in vitro and in vivo). NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug

targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 33; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 33 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of

the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 33; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 33 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 33; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 33 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein.

The product "mature" form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host cell) in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single-stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid

molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or a complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule

that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, that it can hydrogen bond with few or no mismatches to the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

A "fragment" provided herein is defined as a sequence of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.* they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin

and may have a similar or opposite metabolic activity compared to wild type. A "homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be

translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33; or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues that mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described

below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n*

is an integer between 1 and 33. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm

DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See*, *e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See*, *e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* **78**: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an

integer between 1 and 33, thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33; more preferably at least about 70% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33; still more preferably at least about 80% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33; even more preferably at least about 90% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33; and most preferably at least about 95% homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at

one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (*i*) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant NOVX protein and a NOVX ligand; or (*iii*) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Interfering RNA

In one aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA) mediated gene silencing where expression products of a NOVX gene are targeted by specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region. *See, e.g.*, PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304, WO02/16620, and WO02/29858, each incorporated by reference herein in their entirety. Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, *e.g.*, a transcription factor that binds the NOVX gene promoter, a kinase or phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a Drosophila extract, or by transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence. *See, e.g.*, Tuschl, Zamore, Lehmann, Bartel and Sharp (1999), Genes & Dev. 13: 3191-3197, incorporated herein by reference in its entirety. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments using a 24-well tissue culture plate format.

The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the nucleotides in the 3' overhang are deoxyribonucleotides. Using 2'-deoxyribonucleotides in

the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (*e.g.*, a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (*e.g.*, a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize in vivo to generate siRNA constructs for silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and anti-sense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner.

In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, *e.g.*, a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressor™ RNA Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the expression of around 21-nucleotide siRNAs in, *e.g.*, an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression

vector would experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent ribonuclease called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in Drosophila suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is distinct from DICER. RISC uses the sequence encoded by the antisense siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a desired NOVX sequence beginning 50 to100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA sequence is done against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation. See, Elbashir *et al.* 2001 EMBO J. 20(23):6877-88. Hence, consideration should be taken to accommodate SNPs, polymorphisms, allelic variants or species-specific variations when targeting a desired gene.

In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the genome. Typically, one would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition,

expression of two independent genes can be simultaneously knocked down by using equal concentrations of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted NOVX region is typically a sequence of two adenines (AA) and two thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT). A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of the sense siRNA to TT can be performed if such a sequence does not naturally occur in the NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with approximately equal ratios of sense and antisense target RNA-cleaving siRNPs. *See*, *e.g.*, Elbashir, Lendeckel and Tuschl (2001). Genes & Dev. 15: 188-200, incorporated by reference herein in its entirely. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense siRNA strand guides target recognition.

Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21) sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to specificity. Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. *See*, Harborth, *et al.* (2001) J. Cell Science 114: 4557-4565, incorporated by reference in its entirety.

Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (commercially available from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No NOVX gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84 μg of the siRNA duplex is generally sufficient. Cells are typically

seeded the previous day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (*e.g.* inversion versus vortexing) are also critical. Low transfection efficiencies are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal, more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated as within the scope of the present invention.

For a control experiment, transfection of 0.84 μg single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84 μg antisense siRNA has a weak silencing effect when compared to 0.84 μg of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (*e.g.* commercially available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such techniques as immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no NOVX knock-down phenotype is observed, depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a

non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX protein may exist in the cell. Multiple transfection in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a NOVX siRNA construct as therapy to compensate for increased or aberrant NOVX expression or activity. The NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques, as described above. A NOVX siRNA specific for a NOVX gene will decrease or knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues.

The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a double stranded gene transcript that is processed into siRNAs. Upon treatment, the target protein is not produced or is not produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a control sample of cells or tissues from healthy individuals provides a reference standard for determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, ELISA, and the like. A subject sample of cells or tissues is taken from a mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA constructs, that are specific for the NOVX gene product. These cells or tissues are treated by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the assays described. This NOVX gene knockdown approach provides a rapid method for determination of a NOVX minus (NOVX$^-$) phenotype in the treated subject

sample. The NOVX phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and use of a NOVX siRNA are known to those skilled in the art. Example techniques are provided below.

Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The produced ssRNA and asRNA (0.5 μM) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide. See, *e.g.*, Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989).

Lysate Preparation

Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an additional 60 min. The molar ratio of double stranded RNA and mRNA is about 200:1. The NOVX mRNA is radiolabeled (using known techniques) and its stability is monitored by gel electrophoresis.

In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a ^{32}P-ATP. Reactions are stopped by the addition of 2 X proteinase K buffer and deproteinized as described previously (Tuschl *et al.*, Genes Dev., 13:3191-3197 (1999)). Products are analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluded. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer

for each assay. The sequence of these 21-23 mers is then determined using standard nucleic acid sequencing techniques.

RNA Preparation

21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonucleotides are deprotected and gel-purified (Elbashir, Lendeckel, & Tuschl, Genes & Dev. 15, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, Mass., USA) purification (Tuschl, et al., Biochemistry, 32:11658-11668 (1993)).

These RNAs (20 μM) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90° C followed by 1 h at 37° C.

Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics (1-3 X 105 cells/ml) and transferred to 24-well plates (500 ml/well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express NOVX. Base-paired 21 and 22 nt siRNAs with overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be performed using the same siRNA using well known in vivo transfection or gene therapy transfection techniques.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons that are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically

synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to

peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. *See, e.g.*, Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g.*, Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (*See, e.g.*, Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.,* Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.,* Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*See,* Hyrup, *et al.,* 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*See,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.,* 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity.

PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see*, Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or

residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about 10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20%

chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide that is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, and retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the

molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX

fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically-active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically-active portions of a NOVX protein. In yet another embodiment, a NOVX fusion protein comprises at least three biologically-active portions of a NOVX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is a NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional

techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade that includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene

sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods that can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived that encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the

frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g.*, Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 33, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a

hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate that regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant (K_D) is ≤ 1 μM, preferably ≤ 100 nM, more preferably ≤ 10 nM, and most preferably ≤ 100 pM to about 1 pM, as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with

the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants that can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen that is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as

radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding,1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80:

2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals that are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal that provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells that secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to

obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers that are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (*e.g.* tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or

threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.* $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker that is too short to allow pairing

between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm that binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells that express a particular antigen. These antibodies possess an antigen-binding arm and an arm that binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For

example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as gluteraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as

described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al ., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (e.g., a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably

one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand that may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, *et al.*, editors) Mack Pub. Co., Easton, Pa.: 1995; IN: Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends,

Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of

lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", Diamandis and Christopoulus, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-an analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory

sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*,

tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of

such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable

medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be

operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared that contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the

vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it

develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium

hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and

used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser that contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be

prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.,* Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances

associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell that expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell that expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound that binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially

bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell that expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, *etc.*), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound that binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound that binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples

of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target

molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In

the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein that interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human

chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for

marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for

prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 33, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended

to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a

suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the

integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a NOVX gene; (*ii*) an addition of one or more nucleotides to a NOVX gene; (*iii*) a substitution of one or more nucleotides of a NOVX gene, (*iv*) a chromosomal rearrangement of a NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of a NOVX gene, (*vi*) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (*viii*) a non-wild-type level of a NOVX protein, (*ix*) allelic loss of a NOVX gene, and (*x*) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art that can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid

amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.*, Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT

International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.*, Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.*, Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility

between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer

where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages

and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a

marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize

activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see*, *e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a

disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situ*ations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

Example 1.

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	595 bp	
NOV1a, CG116579-01 DNA Sequence	CGCTGCGGAAAGTTGGGGCAACCTGTTGCTAGTCTGGTCGTTGGTGACAGCGAGGCTTCCGCGCTCGC TGCTGGTGAGCAGCCCCGGCGTGCCCCGCGGGCTGGAAGAGGCGGCGGCGTGATGCGGCCCGTGGACG CGGACGAGGCGCGGGAGCCCCGCGAGGAGCCGGGCAGCCCGCTGAGCCCCGCGCCCCGCGCCGGCCGC GAGAACCTGGCCTCCCTGGAGCGCGAGCGCGCCCGGGCGCACTGGCGGGCCCGCAGGAAGCTGCTGGA GATCCAGAGCCTGCTCGACGCCATCAAGAGTGAGGTGGAGGCAGAGGAGCGGGGCGCCCGGGCCCCAG CACCCCGCCCGCGTGCGGAGGCTGAGGAGCGGGTGGCTCGGCTGTGCGCCGAAGCAGAGAGGAAGGCT GCGGAGGCGGCGCGGATGGGCAGGCGGATCGTGGAGCTGCACCAGCGGATCGCCGGCTGCGAGTGCTG CTGAGCCGGCGAGGCCGCGCGGGTCTGGAGCGGAGCGCGGCGGGGAGTGTCCCGCGTGGAAGGCGCTG GGTGGCCAACTGACGAACTGTGTCACCTGATAAGGAGTCGTGCTGCTGGAC		
	ORF Start: at 4		ORF Stop: TGA at 478
	SEQ ID NO: 2	158 aa	MW at 17097.1kD
NOV1a, CG116579-01 Protein Sequence	CGKLGQPVASLVVGDSEASALAAGEQPRRAPRAGRGGGVMRPVDADEAREPREEPGSPLSPAPRAGRE NLASLERERARAHWRARRKLLEIQSLLDAIKSEVEAEERGARAPAPRPRAEAEERVARLCAEAERKAA EAARMGRRIVELHQRIAGCECC		
	SEQ ID NO: 3	717 bp	
NOV1b, CG116579-02 DNA Sequence	ATGCCTGAGCCTCCCACCCCCTCCATGGGCTCCTGTGCTGCCCGAGCCTCCTCGACGAGCACCACCCC CTGCTCCACGGCGCCCAGTCCCATCACCACCCAAGGGCTGAGGAGTGTGGGCGCACGGTGCCGTTCTC GCAGGCAGCTCCACCTGCAGCCCTGGACCGCAGGTGTCGTAAAAGGGCCGCAGTGGCAGCGTCCTGGC CGACGGCTAGTAGCCCATTTTGGATACCGTCCTCGCTGCGGAAAGTTGGGGCAACCTGTTGCTAGTCT GGTCGTTGGTGACAGCGAGGCTTCCGCGCTCGCTGCTGGTGAGCAGCCCCGGCGTGCCCCGCGGGCTG GAAGAGGCGGCGGCGTGATGCGGCCCGTGGACGCGGACGAGGCGCGGGAGCCCCGCGAGGAGCCGGGC AGCCCGCTGAGCCCCGCGCCCCGCGCCGGCCGCGAGAACCTGGCCTCCCTGGAGCGCGAGCGCGCCCG GGCGCACTGGCGGGCCCGCAGGAAGCTGCTGGAGATCCAGAGCCTGCTCGACGCCATCAAGAGTGAGG TGGAGGCAGAGGAGCGGGGCGCCCGGGCCCCAGCACCCCGCCCGCGTGCGGAGGCTGAGGAGCGGGTG GCTCGGCTGTGCGCCGAAGCAGAGAGGAAGGCTGCGGAGGCGGCGCGGATGGGCAGGCGGATCGTGAA GCTGCACCAGCGGATCGCCGGCTGCGAGTGCTGCTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 715
	SEQ ID NO: 4	238 aa	MW at 25761.0kD
NOV1b, CG116579-02 Protein Sequence	MPEPPTPSMGSCAARASSTSTTPCSTAPSPITTQGLRSVGARCRSRRQLHLQPWTAGVVKGPQWQRPG RRLVAHFGYRPRCGKLGQPVASLVVGDSEASALAAGEQPRRAPRAGRGGGVMRPVDADEAREPREEPG SPLSPAPRAGRENLASLERERARAHWRARRKLLEIQSLLDAIKSEVEAEERGARAPAPRPRAEAEERV ARLCAEAERKAAEAARMGRRIVKLHQRIAGCECC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 1B.

Table 1B. Comparison of NOV1a against NOV1b.		
Protein Sequence	**NOV1a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**

NOV1b	1..158 81..238	79/158 (50%) 80/158 (50%)

Further analysis of the NOV1a protein yielded the following properties shown in Table 1C.

Table 1C. Protein Sequence Properties NOV1a

PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1D.

Table 1D. GENESEQ Results for NOV1a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU72789	Human anticancer protein #2 - Homo sapiens, 127 aa. [CN1313298-A, 19-SEP-2001]	40..156 1..125	55/125 (44%) 69/125 (55%)	2e-17
AAU27977	Human contig polypeptide sequence #130 - Homo sapiens, 164 aa. [WO200164834-A2, 07-SEP-2001]	40..156 38..162	55/125 (44%) 69/125 (55%)	2e-17
AAU27805	Human full-length polypeptide sequence #130 - Homo sapiens, 127 aa. [WO200164834-A2, 07-SEP-2001]	40..156 1..125	55/125 (44%) 69/125 (55%)	2e-17
ABP41776	Human ovarian antigen HNOKE42, SEQ ID NO:2908 - Homo sapiens, 148 aa. [WO200200677-A1, 03-JAN-2002]	40..156 22..146	53/125 (42%) 67/125 (53%)	7e-15
AAY73333	HTRM clone 1760185 protein sequence - Homo sapiens, 127 aa. [WO9957144-A2, 11-NOV-1999]	40..156 1..125	53/125 (42%) 67/125 (53%)	7e-15

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1E.

Table 1E. Public BLASTP Results for NOV1a				
Protein Accession Number	**Protein/Organism/Length**	**NOV1a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8WU25	Hypothetical 13.4 kDa protein - Homo sapiens (Human), 119 aa.	40..158 1..119	119/119 (100%) 119/119 (100%)	1e-62
Q96HT8	Unknown (protein for MGC:9651) (PP784) - Homo sapiens (Human), 127 aa.	40..156 1..125	55/125 (44%) 69/125 (55%)	7e-17
Q9Y605	T-cell activation protein - Homo sapiens (Human), 127 aa.	40..156 1..125	53/125 (42%) 67/125 (53%)	2e-14
Q9CQL7	9130413I22Rik protein (RIKEN cDNA 9130413I22 gene) - Mus musculus (Mouse), 125 aa.	40..156 1..123	54/133 (40%) 69/133 (51%)	3e-10
Q9CX68	9130413I22Rik protein - Mus musculus (Mouse), 126 aa.	40..147 1..114	50/124 (40%) 64/124 (51%)	7e-09

PFam analysis predicts that the NOV1a protein contains the domains shown in the Table 1F.

Table 1F. Domain Analysis of NOV1a			
Pfam Domain	**NOV1a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
No Significant Matches Found To Publicly Searchable Domains			

Example 2.

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis		
	SEQ ID NO: 5	3411 bp
NOV2a, CG126119-01 DNA Sequence	NGGGCCTTCTGGTGGACCTCCACTCCCACGCGGGGCGGGGTGCAGGTGGCGAAGTGGAAATGGAGGCG GGAGATGGAACGCCCCCACCCTCCCTCGACACTTTGGGGCCATGAAAACCCATTCTCGGACCTTCCCA GCGGCACCCTCAATTTTCACCCGGTGTGGACATCTCGAACTTGCTCCCGCCCACCCTTCTGTCTCTCT CAAATAGTACAGCTTAAAGCAATAAATGTAGATCTTCAAAGTGATGCTGCTCTGCAGGTGGACATTTC TGATGCTCTTAGTGAGCGGGATAAAGTAAAATTCACTGTTCACACAAAGAGTTCATTGCCAAATTTTA AACAAAACGAGTTTTCAGTTGTTCGGCAACATGAGGAATTTATCTGGCTTCATGATTCCTTTGTTGAA AATGAAGACTATGCAGGTTATATCATTCCACCAGCACCACCAAGACCTGATTTTGATGCTTCAAGGGA AAAACTACAGAAGCTTGGTGAAGGAGAAGGGTCAATGACGAAGGAAGAATTCACAAAGATGAAACAGG AACTGGAAGCTGAATATTTGGCAATATTCAAGAAGACAGTTGCGATGCATGAAGTGTTCCTGTGTCGT GTGGCAGCACATCCTATTTTGAGAAGAGATTTAAATTTCCATGTCTTCTTGGAATATAATCAAGATTT GAGTGTGCGAGGAAAAAATAAAAAAGAGAAACTTGAAGACTTCTTTAAAAACATGGTTAAATCAGCAG	

	ATGGAGTAATCGTTTCAGGAGTAAAGGATGTAGATGATTTCTTTGAGCACGAACGAACATTTCTTTTG GAATATCATAACCGAGTTAAGGATGCATCTGCTAAATCTGATAGAATGACAAGATCCCACAAAAGTGC TGCAGATGATTACAATAGAATTGGTTCTTCATTATATGCTTTAGGAACTCAGGATTCTACAGATATAT GCAAGTTTTTTCTCAAAGTTTCAGAACTGTTCGATAAAACAAGAAAAATAGAAGCACGAGTGTCTGCT GATGAAGACCTCAAACTTTCTGATCTTTTAAAATATTACTTAAGAGAATCTCAAGCTGCTAAGGATCT CCTGTATCGAAGGTCTAGGTCACTAGTGGATTATGAAAATGCTAATAAAGCACTGGATAAAGCAAGAG CAAAAAATAAAGATGTTCTACAGGCCGAAACTTCCCAACAATTATGTTGTCAGAAATTTGAAAAAATA TCTGAGTCTGCAAAACAAGAACTTATAGATTTTAAGACAAGAAGAGTTGCTGCATTCAGAAAAAATTT AGTGGAACTGGCAGAGTTAGAACTGAAGCATGCAAAGGGTAATCTACAGTTGCTGCAGAACTGCCTGG CAGTGTTAAATGGAGACACATAAGCCACACTCCGCCTTCCTGTTAAAAAGGGCTGCCTTCCTTCAAAT TTTATTTTTGTTTTCTTAATGATGTTAAGCATTTATGCTCACTGGAAACAAACAAAAAGCAGCTGAAA AAGTGCATCAACTCCTCTTTTTCTGAGAAACATGGAGCAGCGCACGCCCAGGCGATGCCAGTCTGTGT GCCGTGATGCCGCACTGTGTTCCCCATGACAGTGGTCCATCATCGTGCACTCGTCATACTCAGAAGTC CAAAGTTCATTCTTCTTTAAAGTAGCCTCTATAACTCTGTTTATTTTATAAATAGTATTCCTTATGGC TGCCACTCTTATTTACCTTTAAATAATTTCTGAAATTTAACCTTTTCAGAATGCATTGTTGAAACAAG ATAAAGATTGCCTTTTTTGAATTTTTTAAATTTTGTTTTTAAAAGCATATACCACCTTAGTTCATTCA TGTATCCTGGTAAAGCATCTTAATCAGACTTATTTTTAATTACTGAATATTTCTTAGACGTTTTGGGA CAGATTTTATGTAATCTTTATAAGTATGATTTCTGAAGAAAAGCAAATGCATTAGTATGTTTGCCTTA AACTTGTAGACTAAACCAAGTATTGTAAAATAAACAGCGATAACAGTGATAGTTTTTAACTCTATGGT CATTGTATCACTCTGGAAAATGTGGAGTAGCTGTAATAAATCTACTCCTGTATTATGCTTTACAGTGC AGGTCTTAGTTTTTCTTTTTTTCTCATTTCTTTTGAAATGGCATCTCGAACAAAGTCCACCAATCCCTT TACAAAAGAATGAACTGCTCCTCTGTGTGTACTTCATAGAAGGTGGAATCGGACAGAGGCAGGTTAGT GACAGTTATTCCTGAAATACAGGAGCAGAGTACAGTCTGTTGTGGTTTCCCGGATTCCGCGCCTAGCT CAGCCAATTAAGCATGAGACATAGGCCATTGAGCCACTTAGTAGTTATGCGAGTGGATAGATTGGTAT GTAGAGGGAAAGAGGTCTGCTGTAAAGAACAACACTTGTTTGTCTGTGGGGAAAGAAAAGCAGAATAC TTGAGATGAAAGTTGGCATACAAATAGGATACTATCGCCAGTAGTTATATTACAAACATTATCGGCCT TTCTAGTGTGAATGAACATTAGACACATTATTGTCATTCCTAGTTTAAAGTTAAGGTTGCGTGGTTGG ATTTTTCCACTATCTTTTTCTAATTTTTCTACCATTTGGAGACCGTAGGCATTTGGGCCTGTCACCCC TTGGATGGGTTCCTAGTTTGTTTACATTTTCCTGAACCCTCCTGAGCGCCCGTTCTTGGTCTAATCCC CAGTCGTGATGATTCCACACTTCCTCAGCCGCATGTTGTCTTGCCTCATTCATGAGCTGGTCAGCGTT TCGTCTCTTTAACTGACATGTTCCCCAGTGCTGTTTGAACTGTTGAGTTTCCGTTGCTGGCTGAGTGC GTTTTGTCCTTCACGTAACCTTCGCTGGTAAAAATAAGCCCATGTGATGTCCACCAGTGGATGAATGC TGGACCGAGAGCCCTAGCTTCTGGATCCAGGTCTAGGCCCTTCATCTGCTGCTCTGTGGCCCAGGGCA GGTTTGCTTGACCTCTGCCTCAGTTCTCGACTCTAAAGGACATACTGACCTACCTCACAGGGGTGTTG TGAGGATTAATAAATGTTGGTACTCTGCTTTGGAAATGTGAAAATGCTGTGTAAATGTTAAGAAATAC TAAGTATAGGGCCAGAAGCTATACAGTGTTTCACTTAACCGTTTGCCATTCTGTATTTACCAAGGTGG TCTTTTCTGGGGAAGGAAGTAGAGTGGAAGGTGCATCCCTTGGCCCCTGGTTTACATTATTAGGGTGC TTATTGTAGGAATGCACTCTAAAAAGTGGGCGTAGAATGAAAGCAGCCGTCCAGTGGTCCTCCCTTTT CTGTAGTTTCACTTTTCTTGCTTCAAGTTACAGCAGTCACCTGAAATCTGAAAATACTAAATGAAAAA CTCCAGAAACA
	ORF Start: ATG at 73 | ORF Stop: TAA at 1381
	SEQ ID NO: 6 | 436 aa | MW at 50124.2kD
NOV2a, CG126119-01 Protein Sequence	MERPHPPSTLWGHENPFSDLPSGTLNFHPVWTSRTCSRPPFCLSQIVQLKAINVDLQSDAALQVDISD ALSERDKVKFTVHTKSSLPNFKQNEFSVVRQHEEFIWLHDSFVENEDYAGYIIPPAPPRPDFDASREK LQKLGEGEGSMTKEEFTKMKQELEAEYLAIFKKTVAMHEVFLCRVAAHPILRRDLNFHVFLEYNQDLS VRGKNKKEKLEDFFKNMVKSADGVIVSGVKDVDDFFEHERTFLLEYHNRVKDASAKSDRMTRSHKSAA DDYNRIGSSLYALGTQDSTDICKFFLKVSELFDKTRKIEARVSADEDLKLSDLLKYYLRESQAAKDLL YRRSRSLVDYENANKALDKARAKNKDVLQAETSQQLCCQKFEKISESAKQELIDFKTRRVAAFRKNLV ELAELELKHAKGNLQLLQNCLAVLNGDT
	SEQ ID NO: 7 | 562 bp
NOV2b, CG126119-02 DNA Sequence	GGAGCAGCCATGATGGAAGGCCTGGACGACGGCCCGGACTTCCTCTCAGAAGAGGACCGCGGACTTAA AGCAATAAATGTAGATCTTCAAAGTGATGCTGCTCTGCAGGTGGACATTTCTGATGCTCTTAGTGAGC GGGATAAAGTAAAATTCACTGTTCACACAAAGAGTTCATTGCCAAATTTTAAACAAAACGAGTTTTCA GTTGTTCGGCAACATGAGGAATTTATCTGGCTTCATGATTCCTTTGTTGAAAATGAAGACTATGCAGG TTATATCATTCCACCAGCACCACCAAGACCTGATTTTGATGCTTCAAGGGAAAAACTACAGAAGCTTG GTGAAGGAGAAGGGTCAATGACGAAGGAAGAATTCACAAAGATGAAACAGGAACTGGAAGCGGGTTGG ATAACAGAGAACCTTGGGTTTATTCTACTGCTACCTCCATCCTCTGCATCCTTCTTTTTTGTCTTCAC TGAATGACTACCCTCACAGAGATCAAACTTCTCCCATCATTGGTCCTGCTGGTTTGCTGTGAATATTT GGCAATATTCAAGAAGAA
	ORF Start: ATG at 10 | ORF Stop: TGA at 481
	SEQ ID NO: 8 | 157 aa | MW at 17836.7kD

NOV2b, CG126119-02 Protein Sequence	MMEGLDDGPDFLSEEDRGLKAINVDLQSDAALQVDISDALSERDKVKFTVHTKSSLPNFKQNEFSVVR QHEEFIWLHDSFVENEDYAGYIIPPAPPRPDFDASREKLQKLGEGEGSMTKEEFTKMKQELEAGWITE NLGFILLLPPSSASFFFVFTE

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b.

Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	39..164 9..134	115/126 (91%) 118/126 (93%)

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a

PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV2a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2D.

Table 2D. GENESEQ Results for NOV2a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB43157	Human ORFX ORF2921 polypeptide sequence SEQ ID NO:5842 - Homo sapiens, 460 aa. [WO200058473-A2, 05-OCT-2000]	1..436 25..460	436/436 (100%) 436/436 (100%)	0.0
ABP41711	Human ovarian antigen HPAMC60, SEQ ID NO:2843 - Homo sapiens, 414 aa. [WO200200677-A1, 03-JAN-2002]	39..436 17..414	390/398 (97%) 391/398 (97%)	0.0

AAY94209	Human TRAF four associated factor TFAF2 - Homo sapiens, 406 aa. [CA2245340-A1, 19-FEB-2000]	39..436 9..406	390/398 (97%) 391/398 (97%)	0.0
AAB07856	Amino acid sequence of Smad1 interactor protein clone S12 - Homo sapiens, 414 aa. [WO200047102-A2, 17-AUG-2000]	39..436 17..414	390/398 (97%) 391/398 (97%)	0.0
AAB58368	Lung cancer associated polypeptide sequence SEQ ID 706 - Homo sapiens, 414 aa. [WO200055180-A2, 21-SEP-2000]	39..436 17..414	390/398 (97%) 391/398 (97%)	0.0

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a

Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UNH7	Sorting nexin 6 (TRAF4-associated factor 2) - Homo sapiens (Human), 406 aa.	39..436 9..406	390/398 (97%) 391/398 (97%)	0.0
Q9CZ03	2810425K19Rik protein - Mus musculus (Mouse), 406 aa.	39..436 9..406	388/398 (97%) 391/398 (97%)	0.0
Q9BUY3	Hypothetical 33.6 kDa protein - Homo sapiens (Human), 290 aa.	147..436 1..290	290/290 (100%) 290/290 (100%)	e-163
Q9D8U8	Sorting nexin 5 - Mus musculus (Mouse), 404 aa.	48..429 17..398	254/382 (66%) 323/382 (84%)	e-152
Q9Y5X3	Sorting nexin 5 - Homo sapiens (Human), 404 aa.	48..429 17..398	252/382 (65%) 323/382 (83%)	e-152

PFam analysis predicts that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a

Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PX	59..200	35/160 (22%) 106/160 (66%)	4.9e-19

Example 3.

The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis

	SEQ ID NO: 9	929 bp	
NOV3a, CG137623-01 DNA Sequence	AGGCGCCTCAGCCCGGCCTGGGCGAGCCCTGGGTGCTCCGCCGGGCAGCTCACGGCGCCCCGTATGGC CTGGGGATCCTAAGAGGCCCTGTGACCCCCCTCGCCTGGTCTCCCTCTCACCCCTGGAGGGTTGCCGC AGCTCCGGGGCCCCCGGGCAGGAAGGGCGCACTGGTCGTCCCGGGAGAGGGGTCTGAGCAGAGGGCGG GGTGCAGGCGGAATGGCCCTCGTGCCCTATGAGGAGACCACGGAATTTGGGTTGCAGAAATTCCACAA GCCTCTTGCAACTTTTTCCTTTGCAAACCACACGATCCAGATCCGGCAGGACTGGAGACACCTGGGAG TCGCAGCGGTGGTTTGGGATGCGGCCATCGTTCTTTCCACATACCTGGAGATGGGAGCTGTGGAGCTC AGGGGCCGCTCTGCCGTGGAGCTGGGTGCTGGCACGGGGCTGGTGGGCATAGTGGCTGCCCTGCTGGG TGCTCATGTGACTATCACGGATCGAAAAGTAGCATTAGAATTTCTTAAATCAAACGTTCAAGCCAACT TACCTCCTCATATCCAAACTAAAACTGTTGTTAAGGAGCTGACTTGGGGACAAAATTTGGGGAGTTTT TCTCCTGGAGAATTTGACCTGATACTTGGTGCTGATATCATATATTTAGAAGAAACATTCACAGATCT TCTTCAAACACTGGAACATCTCTGTAGCAATCACTCTGTGATTCTTTTAGCATGCCGAATTCGCTATG AACGGGATAACAACTTCTTAGCAATGCTGGAGAGGCAATTTATTGTGAGAAAGGTTCACTACGATCCT GAAAAAGATGTACATATTTACGAAGCACAGAAGAGAAACCAGAAGGAGGACTTATAATTGGCTATAAT TTATAAGAATGTTGTCATTGAGTGTGTCACTTAAGGTCTTAGACT		
	ORF Start: ATG at 217		ORF Stop: TAA at 871
	SEQ ID NO: 10	218 aa	MW at 24612.0kD
NOV3a, CG137623-01 Protein Sequence	MALVPYEETTEFGLQKFHKPLATFSFANHTIQIRQDWRHLGVAAVVWDAAIVLSTYLEMGAVELRGRS AVELGAGTGLVGIVAALLGAHVTITDRKVALEFLKSNVQANLPPHIQTKTVVKELTWGQNLGSFSPGE FDLILGADIIYLEETFTDLLQTLEHLCSNHSVILLACRIRYERDNNFLAMLERQFIVRKVHYDPEKDV HIYEAQKRNQKEDL		
	SEQ ID NO: 11	420 bp	
NOV3b, CG137623-02 DNA Sequence	GGAAAGCGGAGCGCGCGCTCCACGCGGGACCGCCTCCCGGGCCGTCTGAGCAGAGGGCGGGGTGCAGG CGGAATGGCCCTCGTGCCCTATGAGGAGACCACGGAATTTGGGTTGCAGAAATTCCACAAGCCTCTTG CAACTTTTTCCTTTGCAAACCACACGATCCAGATCCGGCAGGACTGGAGACACCTGGGAGTCGCAGCG GTGGTTTGGGATGCGGCCATCGTTCTTTCCACATACCTGGAGATGGGAGCTGTGGAGCTCAGGGGCCG CTCTGCCGTGGAGCTGGGTGCTGGCACGGGGCTGGTGGGCATAGTGGCTGCCCTGCTGGGAGGTGGAA TTTAATTCTCCTCCCCTTGAATATGGGCTGGACAAAGAGAAAAATGGTAGCTCAACAGTGGAGACACC TGGACAGCACTT		
	ORF Start: ATG at 73		ORF Stop: TAA at 343
	SEQ ID NO: 12	90 aa	MW at 9609.0kD
NOV3b, CG137623-02 Protein Sequence	MALVPYEETTEFGLQKFHKPLATFSFANHTIQIRQDWRHLGVAAVVWDAAIVLSTYLEMGAVELRGRS AVELGAGTGLVGIVAALLGGGI		
	SEQ ID NO: 13	743 bp	
NOV3c, CG137623-03 DNA Sequence	GGAGAGGGGTCTGAGCAGAGGGCGGGGTGCAGGCGGAATGGCCCTCGTGCCCTATGAGGAGACCACGG AATTTGGGTTGCAGAAATTCCACAAGCCTCTTGCAACTTTTTCCTTTGCAAACCACACGATCCAGATC CGGCAGGACTGGAGACACCTGGGAGTCGCAGCGGTGGTTTGGGATGCGGCCATCGTTCTTTCCACATA CCTGGAGATGGGAGCTGTGGAGCCCAGGGGCCGCTCTGCCGTGGAGCTGGGTGCTGGCACGGGGCTGG TGGGCATAGTGGCTGCCCTGCTGGGTGCTCATGTGACTATCACGGATCGAAAAGTAGCATTAGAATTT CTTAAATCAAACGTTCAAGCCAACTTACCTCCTCATATCCAAACTAAAACTGTTGTTAAGGAGCTGAC CTGGGGACAAAATTTGGGGAGTTTTTCTCCTGGAGAATTTGACCTGATACTTGGTGCTGATATCATAT ATTTAGAAGAAACATTCACAGATCTTCTTCAAACACTGGAACATCTCTGTAGCAATCACTCTGTGATT CTTTTAGCATGCCGAATTCGCTATGAACGGGATAACAACTTCTTAGCAATGCTGGAGAGGCAATTTAT TGTGAGAAAGGTTCACTACGATCCTGAAAAAGATGTACATATTTACGAAGCACAGAAGAGAAACCAGA AGGAGGACTTATAATTGGCTATAATTTATAAGAATGTTGTCATTGAGTGTGTCACTTAAGGTC		
	ORF Start: ATG at 38		ORF Stop: TAA at 692
	SEQ ID NO: 14	218 aa	MW at 24596.0kD

NOV3c, CG137623-03 Protein Sequence	MALVPYEETTEFGLQKFHKPLATFSFANHTIQIRQDWRHLGVAAVVWDAAIVLSTYLEMGAVEPRGRS AVELGAGTGLVGIVAALLGAHVTITDRKVALEFLKSNVQANLPPHIQTKTVVKELTWGQNLGSFSPGE FDLILGADIIYLEETFTDLLQTLEHLCSNHSVILLACRIRYERDNNFLAMLERQFIVRKVHYDPEKDV HIYEAQKRNQKEDL

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b and NOV3c.

Protein Sequence	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV3b	1..71 1..71	71/71 (100%) 71/71 (100%)
NOV3c	1..218 1..218	217/218 (99%) 217/218 (99%)

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a

PSort analysis:	0.8500 probability located in endoplasmic reticulum (membrane); 0.7900 probability located in plasma membrane; 0.3520 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV3a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

Table 3D. GENESEQ Results for NOV3a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41517	Human polypeptide SEQ ID NO 6448 - Homo sapiens, 161 aa. [WO200153312-A1, 26-JUL-2001]	1..90 29..118	86/90 (95%) 87/90 (96%)	3e-43
AAM39731	Human polypeptide SEQ ID NO 2876 - Homo sapiens, 133 aa. [WO200153312-A1, 26-JUL-2001]	1..90 1..90	86/90 (95%) 87/90 (96%)	3e-43

ABB80681	Human transferase protein, 26199 - Homo sapiens, 229 aa. [WO200220801-A2, 14-MAR-2002]	30..215 26..226	77/204 (37%) 108/204 (52%)	4e-23
AAM40002	Human polypeptide SEQ ID NO 3147 - Homo sapiens, 2505 aa. [WO200153312-A1, 26-JUL-2001]	32..96 1106..1170	48/65 (73%) 55/65 (83%)	3e-19
AAG27905	Arabidopsis thaliana protein fragment SEQ ID NO: 32925 - Arabidopsis thaliana, 325 aa. [EP1033405-A2, 06-SEP-2000]	23..176 15..183	54/170 (31%) 82/170 (47%)	9e-17

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

Table 3E. Public BLASTP Results for NOV3a

Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8WXB1	Hepatocellular carcinoma-associated antigen HCA557b - Homo sapiens (Human), 218 aa.	1..218 1..218	218/218 (100%) 218/218 (100%)	e-123
BAC04229	CDNA FLJ36493 fis, clone THYMU2018547 - Homo sapiens (Human), 218 aa.	1..218 1..218	217/218 (99%) 217/218 (99%)	e-122
Q9CQL0	2310038H17Rik protein - Mus musculus (Mouse), 218 aa.	1..218 1..218	195/218 (89%) 204/218 (93%)	e-109
Q8R2Y7	RIKEN cDNA 2310038H17 gene - Mus musculus (Mouse), 218 aa.	1..218 1..218	194/218 (88%) 203/218 (92%)	e-108
Q95K98	Hypothetical 18.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 163 aa.	47..218 7..163	155/172 (90%) 155/172 (90%)	6e-82

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3F.

Table 3F. Domain Analysis of NOV3a

Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found To Publicly Searchable Domains			

Example 4.

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis

	SEQ ID NO: 15	1056 bp	
NOV4a, CG137687-01 DNA Sequence	CCGGTGCCGGCCGCGCCATTGTTGGGGGAGGGGGCGGCTGTTGAGGTCGGCGGAGTAGGGGGCGAGCG AAGGCGGTGGCAGAGAGGAGCGGAGGCTTCCCATGGGGAACACGCTGACCTGTTGCGTGTCCCCCAAT GCCAGCCCCAAGCTGGGCCGGCGCGCGGGGTCGGGGAGCTGTACTGCGCGTCCGACATCTACGAGGCG GTGTCCGGGAGAGGGCGAGGGCCACCACCTGCAGCACATCAGCGACCGCGAGATGCCCGAAGATTTAG CTTTGGAGTCAAACCCTTCTGACCATCCAAGGGCAAGCACAATTTTCCTGAGCAAATCTCAAACGGAT GTGCGAGAAAAGAGGAAGAGCAACCATTTGAACCATGTAAGTCCAGGGCAGCTTACTAAAAAGTATAG CTCATGCTCAACAATATTTCTAGATGACAGCACAGTCAGCCAGCCTAATCTTAGAACCACAGTAAAAA GTGTGACCTTAGCAATATATTACCACATAAAGAACAGGCTTCAAGATGCAAATAGATCCCTGGATATT TTTGATGAGAGATCACATCCACTTACAGTAAGTAGTGAGCCAGTTTACTTAGAAAGGCTTTTAACTTA TGCTGAAATCGACATTTGTCCCACCAACTGGAAAAGGATTGTTCTGGGAGCCATTCTTCTTGCCTCCA AGGTTTGGGACGATCAGGCTGTATGGAATGTGGACTACTGCCAGATCCTCAAGGACATTACAGTTGAG GACATGAATGAAATGGAAAGGCATTTTCTGGAGCTTCTTCAGTTTAATATTAATGTTCCTGCCAGTGT TTATGCCAAATACTACTTTGACCTTCGCTCCTTAGCAGATGACAACAACCTGAATTTTCTATTTGCTC CTCTTAGCAAAGAAAGAGCACAGAACCTAGAGGCTATTTCTAGATTGTGTGAAGACAAAGACTTGTGT AGAGCCGCTATGAGAAGGTCTTTCAGTGCTGATAACTTCATTGGTATTCAGCGCTCTAAAGCCATCCT CTCTTAAAAGGAGAAATGAGGGGTTATAACGTCATG		
	ORF Start: ATG at 101		ORF Stop: TAA at 1025
	SEQ ID NO: 16	308 aa	MW at 34869.9kD
NOV4a, CG137687-01 Protein Sequence	MGNTLTCCVSPNASPKLGRRAGSGSCTARPTSTRRCPGEGEGHHLQHISDREMPEDLALESNPSDHPR ASTIFLSKSQTDVREKRKSNHLNHVSPGQLTKKYSSCSTIFLDDSTVSQPNLRTTVKSVTLAIYYHIK NRLQDANRSLDIFDERSHPLTVSSEPVYLERLLTYAEIDICPTNWKRIVLGAILLASKVWDDQAVWNV DYCQILKDITVEDMNEMERHFLELLQFNINVPASVYAKYYFDLRSLADDNNLNFLFAPLSKERAQNLE AISRLCEDKDLCRAAMRRSFSADNFIGIQRSKAILS		
	SEQ ID NO: 17	1158 bp	
NOV4b, CG137687-02 DNA Sequence	GCTGTTGAGGTCGGCGGAGTAGGGGGCGAGCGAAGGCGGTGGCAGAGAGGAGCGGAGGCTTCCCATGG GGAACACGCTGACCTGTTGCGTGTCCCCCAATGCCAGCCCCAAGCTGGGCCGGCGCGCGGGGTCGGCG GAGCTGTACTGCGCGTCCGACATCTACGAGGCGGTGTCCGGGGACGCGGTGGCGGTAGCGCCCGCTGT GGTGGAGCCTGCCGAGTTGGATTTCGGAGAGGGCGAGGGCCACCACCTGCAGCACATCAGCGACCGCG AGATGCCCGAAGATTTAGCTTTGGAGTCAAACCCTTCTGACCATCCAAGGGCAAGCACAATTTTCCTG AGCAAATCTCAAACGGATGTGCGAGAAAAGAGGAAGAGCAACCATTTGAACCATGTATCTCCAGGGCA GCTTACTAAAAAGTATAGCTCATGCTCAACAATATTTCTAGATGACAGCACAGTCAGCCAGCCTAATC TTAGAACCACAGTAAAATGTGTGACCTTAGCAATATATTACCACATAAAGGACAGAGATGCAAATAGA TCCCTGGATATTTTTGATGAGAGATCACATCCACTTACACGAGAAAAAGTTCCAGAGGAATACTTTAA GCATGATCCTGAGCACAAATTTATTTACAGATTTGTTCGTACTCTTTTTAGTGCTGCACAGCTAACAG CTGAATGTGCAATAGTAACTTTGGTTTACTTAGAAAGGCTTTTAACTTATGCTGAAATCGACATTTGT CCCACCAACTGGAAAAGGATTGTTCTGGGAGCCATTCTTCTTGCCTCCAAGGTTTGGGACGATCAGGC TGTATGGAATGTGGACTACTGCCAGATCCTCAAGGACATTACAGTTGAGGACATGAATGAAATGGAAA GGCATTTTCTGGAGCTTCTTCAGTTTAATATTAATGTTCCTGCCAGTGTTTATGCCAAATACTACTTT GACCTTCGCTCCTTAGCAGATGACAACAACCTGAATTTTCTATTTGCTCCTCTTAGCAAAGAAAGAGT ACAGAACCTAGAGGCTATTTCTAGATTGTGTGAAGACAAAGACTTGTGTAGAGCCGCTATGAGAAGGT CTTTCAGTGCTGATAACTTCATTGGTATTCAGCGCTTTAAAGCCATCCTCTCTTAAAAGGAGAAATGA GG		
	ORF Start: ATG at 65		ORF Stop: TAA at 1142
	SEQ ID NO: 18	359 aa	MW at 40793.7kD
NOV4b, CG137687-02 Protein Sequence	MGNTLTCCVSPNASPKLGRRAGSAELYCASDIYEAVSGDAVAVAPAVVEPAELDFGEGEGHHLQHISD REMPEDLALESNPSDHPRASTIFLSKSQTDVREKRKSNHLNHVSPGQLTKKYSSCSTIFLDDSTVSQP NLRTTVKCVTLAIYYHIKDRDANRSLDIFDERSHPLTREKVPEEYFKHDPEHKFIYRFVRTLFSAAQL TAECAIVTLVYLERLLTYAEIDICPTNWKRIVLGAILLASKVWDDQAVWNVDYCQILKDITVEDMNEM ERHFLELLQFNINVPASVYAKYYFDLRSLADDNNLNFLFAPLSKERVQNLEAISRLCEDKDLCRAAMR RSFSADNFIGIQRFKAILS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 4B.

Table 4B. Comparison of NOV4a against NOV4b.

Protein Sequence	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV4b	1..308 1..359	284/361 (78%) 285/361 (78%)

Further analysis of the NOV4a protein yielded the following properties shown in Table 4C.

Table 4C. Protein Sequence Properties NOV4a

PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV4a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4D.

Table 4D. GENESEQ Results for NOV4a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM40020	Human polypeptide SEQ ID NO 3165 - Homo sapiens, 341 aa. [WO200153312-A1, 26-JUL-2001]	1..308 1..341	226/345 (65%) 248/345 (71%)	e-114
AAM41806	Human polypeptide SEQ ID NO 6737 - Homo sapiens, 352 aa. [WO200153312-A1, 26-JUL-2001]	6..308 17..352	222/340 (65%) 244/340 (71%)	e-112
ABB12231	Human novel protein, SEQ ID NO:2601 - Homo sapiens, 352 aa. [WO200157188-A2, 09-AUG-2001]	6..308 17..352	222/340 (65%) 244/340 (71%)	e-112
ABB90124	Human polypeptide SEQ ID NO 2500 - Homo sapiens, 287 aa. [WO200190304-A2, 29-NOV-2001]	59..308 1..287	197/287 (68%) 213/287 (73%)	e-102

AAM85252	Human immune/haematopoietic antigen SEQ ID NO:12845 - Homo sapiens, 159 aa. [WO200157182-A2, 09-AUG-2001]	158..308 7..159	122/153 (79%) 131/153 (84%)	7e-62

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a

Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAC05160	CDNA FLJ40432 fis, clone TESTI2039227 - Homo sapiens (Human), 289 aa.	53..308 1..289	249/291 (85%) 249/291 (85%)	e-135
CAD39020	Hypothetical protein - Homo sapiens (Human), 353 aa (fragment).	1..308 13..353	226/345 (65%) 248/345 (71%)	e-114
Q8TEX2	Cyclin fold protein 1 variant b - Homo sapiens (Human), 341 aa.	1..308 1..341	226/345 (65%) 248/345 (71%)	e-114
Q95LR5	Hypothetical 34.9 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 301 aa.	57..308 13..301	197/289 (68%) 214/289 (73%)	e-101
Q95LK3	Hypothetical 23.5 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 202 aa.	146..308 3..202	131/200 (65%) 141/200 (70%)	4e-62

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a

Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cyclin	71..234	25/179 (14%) 112/179 (63%)	0.00074

Example 5.

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 19	1308 bp	
NOV5a, CG143198-01 DNA Sequence	ATGGCGGCTAGTGATACAGAGCGAGATGGACTAGCCCCAGAAAAGACATCACCAGATAGAGATAAGAA AAAAGAGCAGTCAGAAGTATCTGTTTCTCCTAGAGCTTCAAAACATCATTATTCAAGATCACGATCAA GGTCAAGAGAAAGAAAACGAAAGTCAGATAATGAAGGAAGAAAACACAGGAGCCGGAGCAGAAGCAAA GAGGGAAGAAGACATGAATCCAAAGATAAATCCTCTAAGAAACATAAGTCTGAGGAACATAATGACAA AGAACATTCTTCTGATAAAGGAAGAGAGCGACTAAATTCATCTGAAAATGGTGAGGACAGGCACAAAC GCAAAGAAAGAAAGTCATCAAGAGGCAGAAGTCACTCAAGATCTAGGTCTCGTGAAAGACGCCATCGT AGTAGAAGCAGGGAGCGGAAGAAGTCTCGATCCAGGAGTAGGGAGCGGAAGAAATCGAGATCCAGAAG CAGAGAGAGGAAGAAATCGAGATCCAGAAGCAGGGAAAGAAAACGGCGGATCAGGTCTCGTTCCCGCT CAAGATCAAGACACAGGCATAGGACTAGAAGCAGGAGTAGGACAAGGAGTAGGAGTCGAGATAGAAAG AAGAGAATTGAAAAGCCGAGAAGATTTAGCAGAAGTTTAAGCCGGACTCCAAGTCCACCTCCCTTCAG AGGCAGAAACACAGCAATGGATGCACAGGAAGCTTTAGCTAGAAGGTTGGAAAGGGCAAAGAAATTAC AAGAACAGCGAGAAAAGGAAATGGTTGAAAAACAAAAACAACAAGAAATAGCTGCAGCAGCTGCAGCT ACTGGAGGTTCTGTTCTCAATGTTGCTGCCCTGTTGGCATCAGGAACACAAGTAACACCTCAGATAGC CATGGCAGCTCAGATGGCAGCCCTGCAAGCTAAAGCTTTGGCAGAGACAGGAATAGCTGTTCCTAGCT ACTATAACCCAGCCGCTGTTAATCCAATGAAATTTGCTGAACAAGAGAAAAAAAGGAAAATGCTTTGG CAGGGCAAGAAAGAAGGGGACAAATCCCAATCTGCTGAAATATGGGAAAAATTGAATTTTGGAAACAA GGACCAAAATGTCAAATTTAGGAAATTGATGGGTATTAAGAGTGAAGATGAAGCTGGATGTAGCTCAG TTGATGAAGAAAGTTACAAGACTCTGAAGCAGCAGGAAGAAGTATTTCGAAATTTAGATGCTCAGTAT GAAATGGCAAGATCACAAACCCACACACAAAGAGGAATGGGTTTGGGTTTCACATCTTCAATGCGAGG AATGGATGCAGTTTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1306
	SEQ ID NO: 20	435 aa	MW at 50630.2kD
NOV5a, CG143198-01 Protein Sequence	MAASDTERDGLAPEKTSPDRDKKKEQSEVSVSPRASKHHYSRSRSRSRERKRKSDNEGRKHRSRSRSK EGRRHESKDKSSKKHKSEEHNDKEHSSDKGRERLNSSENGEDRHKRKERKSSRGRSHSRSRSRERRHR SRSRERKKSRSRSRERKKSRSRSRERKKSRSRSRERKRRIRSRSRSRSRHRHRTRSRSRTRSRSRDRK KRIEKPRRFSRSLSRTPSPPPFRGRNTAMDAQEALARRLERAKKLQEQREKEMVEKQKQQEIAAAAAA TGGSVLNVAALLASGTQVTPQIAMAAQMAALQAKALAETGIAVPSYYNPAAVNPMKFAEQEKKRKMLW QGKKEGDKSQSAEIWEKLNFGNKDQNVKFRKLMGIKSEDEAGCSSVDEESYKTLKQQEEVFRNLDAQY EMARSQTHTQRGMGLGFTSSMRGMDAV		

Further analysis of the NOV5a protein yielded the following properties shown in Table 5B.

Table 5B. Protein Sequence Properties NOV5a	
PSort analysis:	0.9920 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV5a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5C.

Table 5C. GENESEQ Results for NOV5a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV5a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**

AAG67014	Human sperm-specific protein EM1, EM6-48 - Homo sapiens, 435 aa. [WO200168685-A1, 20-SEP-2001]	1..435 1..435	435/435 (100%) 435/435 (100%)	0.0
AAM78680	Human protein SEQ ID NO 1342 - Homo sapiens, 435 aa. [WO200157190-A2, 09-AUG-2001]	1..435 1..435	435/435 (100%) 435/435 (100%)	0.0
AAG74995	Human colon cancer antigen protein SEQ ID NO:5759 - Homo sapiens, 431 aa. [WO200122920-A2, 05-APR-2001]	10..435 6..431	421/426 (98%) 422/426 (98%)	0.0
AAG62626	Human RNA helicase 43 - Homo sapiens, 387 aa. [WO200138368-A1, 31-MAY-2001]	53..435 2..387	370/386 (95%) 375/386 (96%)	0.0
AAM79664	Human protein SEQ ID NO 3310 - Homo sapiens, 399 aa. [WO200157190-A2, 09-AUG-2001]	1..352 42..393	352/352 (100%) 352/352 (100%)	0.0

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5D.

Table 5D. Public BLASTP Results for NOV5a

Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9P068	HSPC314 - Homo sapiens (Human), 248 aa (fragment).	137..369 15..248	218/234 (93%) 222/234 (94%)	e-115
Q9CW29	1500011J06Rik protein - Mus musculus (Mouse), 255 aa (fragment).	69..278 18..227	202/210 (96%) 210/210 (99%)	e-110
Q9H864	CDNA FLJ13923 fis, clone Y79AA1000539 (Hypothetical 22.4 kDa protein) - Homo sapiens (Human), 202 aa.	233..435 1..202	202/203 (99%) 202/203 (99%)	e-108
Q9CSJ3	1500011J06Rik protein - Mus musculus (Mouse), 194 aa (fragment).	68..253 9..194	179/186 (96%) 186/186 (99%)	3e-98
Q95TP3	LD33732p - Drosophila melanogaster (Fruit fly), 492 aa.	2..419 33..471	125/465 (26%) 204/465 (42%)	1e-30

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5E.

Table 5E. Domain Analysis of NOV5a			
Pfam Domain	**NOV5a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
No Significant Matches Found To Publicly Searchable Domains			

Example 6.

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis			
	SEQ ID NO: 21	579 bp	
NOV6a, CG144756-01 DNA Sequence	GGAGGAACACGAGACTGAGAGATGAATATTCAACAGAGGCTGCAAAGCCTGTGGACTTTAGCCAGACCCT TCTGCCCTCCTTTGCTGGCGACAGCCTCTCAAATGCAGATGGTTGTGCTCCCTTGCCTGGGTTTTACCCT GCTTCTCTGGAGCCAGGTATCAGGGGCCCAGGGCCAAGAATTCCACTTTGGGCCCTGCCAAGTGAAGGGG GTTGTTCCCCAGAAACTGTGGGAAGCCTTCTGGGCTGTGAAAGACACTATGCAAGCTCAGGATAACATCA CGAGTGCCCGGCTGCTGCAGCAGGAGGTTCTGCAGAACGTCTCGCAAGAAAATGAGATGTTTTCCATCAG AGACAGTGCACACAGGCGGTTTCTGCTATTCCGGAGAGCATTCAAACAGTTGGACGTAGAAGCAGCTCTG ACCAAAGCCCTTGGGGAAGTGGACATTCTTCTGACCTGGATGCAGAAATTCTACAAGCTCTGAATGTCTA GACCAGGACCTCCCTCCCCCTGGCACTGGTTTGTTCCCTGTGTCATTTCAAACAGTCTCCCTTCCTATGC TGTTCACTGGACACTTCAC		
	ORF Start: ATG at 22		ORF Stop: TGA at 481
	SEQ ID NO: 22	153 aa	MW at 17629.5kD
NOV6a, CG144756-01 Protein Sequence	MNIQQRLQSLWTLARPFCPPLLATASQMQMVVLPCLGFTLLLWSQVSGAQGQEFHFGPCQVKGVVPQKLW EAFWAVKDTMQAQDNITSARLLQQEVLQNVSQENEMFSIRDSAHRRFLLFRRAFKQLDVEAALTKALGEV DILLTWMQKFYKL		
	SEQ ID NO: 23	740 bp	
NOV6b, CG144756-02 DNA Sequence	GTGAGGAACACGAGACTGAGAGATGAATTTTCAACAGAGGCTGCAAAGCCTGTGGACTTTAGCCAGACCC TTCTGCCCTCCTTTGCTGGCGACAGCCTCTCAAATGCAGATGGTTGTGCTCCCTTGCCTGGGTTTTACCC TGCTTCTCTGGAGCCAGGTATCAGGGGCCCAGGGCCAAGAATTCCACTTTGGGCCCTGCCAAGTGAAGGG GGTTGTTCCCCAGAAACTGTGGGAAGCCTTCTGGGCTGTGAAAGACACTATGCAAGCTCAGGATAACATC ACGAGTGCCCGGCTGCTGCAGCAGGAGGTTCTGCAGAACGTCTCGGATGCTGAGAGCTGTTACCTTGTCC ACACCCTGCTGGAGTTCTACTTGAAAACTGTTTTCAAAAACTACCACAATAGAACAGTTGAAGTCAGGAC TCTGAAGTCATTCTCTACTCTGGCCAACAACTTTGTTCTCATCGTGTCACAACTGCAACCCAGTCAAGAA AATGAGATGTTTTCCATCAGAGACAGTGCACACAGGCGGTTTCTGCTATTCCGGAGAGCATTCAAACAGT TGGACGTAGAAGCAGCTCTGACCAAAGCCCTTGGGGAAGTGGACATTCTTCTGACCTGGATGCAGAAATT CTACAAGCTCTGAATGTCTAGACCAGGACCTCCCTCCCCCTGGCACTGGTTTGTTCCCTGTGTCATTTCA AACAGTCTCCCTTCCTATGCTGTTCACTGGACACTTCATA		
	ORF Start: ATG at 23		ORF Stop: TGA at 641
	SEQ ID NO: 24	206 aa	MW at 23824.5kD
NOV6b, CG144756-02 Protein Sequence	MNFQQRLQSLWTLARPFCPPLLATASQMQMVVLPCLGFTLLLWSQVSGAQGQEFHFGPCQVKGVVPQKLW EAFWAVKDTMQAQDNITSARLLQQEVLQNVSDAESCYLVHTLLEFYLKTVFKNYHNRTVEVRTLKSFSTL ANNFVLIVSQLQPSQENEMFSIRDSAHRRFLLFRRAFKQLDVEAALTKALGEVDILLTWMQKFYKL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b.

Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..153 1..206	136/206 (66%) 138/206 (66%)

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

Table 6C. Protein Sequence Properties NOV6a

PSort analysis:	0.7480 probability located in microbody (peroxisome); 0.6500 probability located in plasma membrane; 0.3000 probability located in Golgi body; 0.2100 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 50 and 51

A search of the NOV6a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

Table 6D. GENESEQ Results for NOV6a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE07313	Human cancer-specific Mob-5 (cMob-5) mutant protein - Homo sapiens, 154 aa. [WO200155170-A1, 02-AUG-2001]	1..153 1..154	152/154 (98%) 152/154 (98%)	2e-83
AAB65295	Human PRO3301 protein sequence SEQ ID NO:507 - Homo sapiens, 206 aa. [WO200073454-A1, 07-DEC-2000]	1..153 1..206	152/206 (73%) 152/206 (73%)	3e-77
AAB35268	Human mda-7 protein - Homo sapiens, 206 aa. [WO200105437-A2, 25-JAN-2001]	1..153 1..206	152/206 (73%) 152/206 (73%)	3e-77
AAU29220	Human PRO polypeptide sequence #197 - Homo sapiens, 206 aa. [WO200168848-A2, 20-SEP-2001]	1..153 1..206	152/206 (73%) 152/206 (73%)	3e-77

AAY42304	Human tumour suppressor protein mda-7 - Homo sapiens, 206 aa. [WO9947709-A2, 23-SEP-1999]	1..153 1..206	152/206 (73%) 152/206 (73%)	3e-77

In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6E.

Table 6E. Public BLASTP Results for NOV6a

Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96KG4	Suppression of tumorigenicity 16 protein - Homo sapiens (Human), 206 aa.	1..153 1..206	152/206 (73%) 152/206 (73%)	9e-77
Q13007	Interleukin-24 precursor (Suppression of tumorigenicity 16 protein) (Melanoma differentiation associated protein 7) (MDA-7) - Homo sapiens (Human), 206 aa.	1..153 1..206	152/206 (73%) 152/206 (73%)	9e-77
Q925J3	Th2-specific cytokine FISP - Mus musculus (Mouse), 220 aa.	21..153 35..220	87/186 (46%) 103/186 (54%)	2e-37
Q925S4	Melanoma differentiation associated gene-7 - Mus musculus (Mouse), 181 aa.	32..153 7..181	84/175 (48%) 98/175 (56%)	3e-36
Q9WVP8	C49a - Rattus norvegicus (Rat), 183 aa.	26..153 3..183	80/181 (44%) 97/181 (53%)	6e-33

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a

Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found To Publicly Searchable Domains			

Example 7.

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis			
	SEQ ID NO: 25	1210 bp	
NOV7a, CG145473-01 DNA Sequence	GAGAGACACTCCCGAGCGCCGTAAATAGAGTCCAAGTGGGCGGAGAGCCGTCCCGCGCCGCCCGCTCA TGTCTCTACAGAGCCGACTGTCCGGCCGCCTGGCACAGCTGCGCGCGGCGGGGCAGCTGCTCGTCCCC CCGCGCCCCCGGCCCGGACACTTGGCGGGTGCCACGAGGACCCGCAGCAGCACGTGCGGTCCCCCGGC GTTCCTGGGCGTGTTCGGCCGCCGTGCGCGGACCTCGGCGGGAGTTGGGGCGTGGGGGGCGGCGGCGG TGGGGCGGACAGCCGGGGTGCGCACTTGGGCCCCCCTGGCCATGGCGGCGAAGGTGGACCTGAGCACC TCCACCGACTGGAAGGAGGCGAAATCCTTTCTGAAGGGCCTGAGTGACAAGCAGCGGGAGGAACATTA CTTCTGCAAGGACTTTGTCAGGCTGAAGAAGATCCCGACATGGAAGGAGATGGCGAAAGGGGTGGCTG TGAAGGTGGAGGAGCCCAGGTATAAAAAGGACAAGCAGCTCAATGAGAAAATCTCCCTGCTCCGCAGC GACATCACCAAGCTGGAGGTGGACGCCATCGTCAACGCCGCCAACAGCTCCCTGCTCGGAGGCGGTGG CGTGGACGGCTGCATTCATCGGGCCGCCGGCCCCCTGCTTACCGACGAGTGCCGGACCCTGCAGAGCT GTAAGACTGGCAAGGCCAAGATCACCGGCGGCTATCGGCTCCCGGCCAAGTACGTCATCCACACAGTG GGGCCCATCGCCTACGGGGAGCCCAGTGCCAGCCAGGCTGCCGAGCTCCGCAGCTGCTACCTGAGCAG TCTGGACCTGCTGCTGGAGCACCGGCTCCGCTCGGTGGCGTTCCCCTGCATCTCCACCGGCGTGTTTG GCTACCCCTGTGAGGCGGCCGCCGAGATCGTGCTGGCCACGCTGCGAGAGTGGCTGGAGCAGCACAAG GACAAGGACGAGGACATCTACCGGAGCCGGCTCCCCCACTACTTCCCCGTGGCCTGAGGCTCCCGCAG CCCACCCTGACCGGGACTGACCCGCCTTCGGGACCCCGCTCCCAGCTCTGAGAGGTCGCCAAAGCCTG CAGCCTGGCCTGGGCCTGGCCACCCCTTCTTTCCCTCCGCGCCCCGCCCCCGAGGAGCCTAATAAAGA TCTCGTTGTGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 68		ORF Stop: TGA at 1007
	SEQ ID NO: 26	313 aa	MW at 34074.9kD
NOV7a, CG145473-01 Protein Sequence	MSLQSRLSGRLAQLRAAGQLLVPPRPRPGHLAGATRTRSSTCGPPAFLGVFGRRARTSAGVGAWGAAA VGRTAGVRTWAPLAMAAKVDLSTSTDWKEAKSFLKGLSDKQREEHYFCKDFVRLKKIPTWKEMAKGVA VKVEEPRYKKDKQLNEKISLLRSDITKLEVDAIVNAANSSLLGGGGVDGCIHRAAGPLLTDECRTLQS CKTGKAKITGGYRLPAKYVIHTVGPIAYGEPSASQAAELRSCYLSSLDLLLEHRLRSVAFPCISTGVF GYPCEAAAEIVLATLREWLEQHKDKDEDIYRSRLPHYFPVA		

Further analysis of the NOV7a protein yielded the following properties shown in Table 7B.

Table 7B. Protein Sequence Properties NOV7a	
PSort analysis:	0.6106 probability located in mitochondrial matrix space; 0.5199 probability located in mitochondrial intermembrane space; 0.3430 probability located in microbody (peroxisome); 0.3142 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 19 and 20

A search of the NOV7a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 7C.

Table 7C. GENESEQ Results for NOV7a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV7a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAY58613	Protein regulating gene expression	83..291 1..209	209/209 (100%) 209/209 (100%)	e-118

	[WO9964596-A2, 16-DEC-1999]			
AAB58880	Breast and ovarian cancer associated antigen protein sequence SEQ ID 588 - Homo sapiens, 133 aa. [WO200055173-A1, 21-SEP-2000]	67..176 3..112	110/110 (100%) 110/110 (100%)	3e-57
ABB49353	Listeria monocytogenes protein #2057 - Listeria monocytogenes, 176 aa. [WO200177335-A2, 18-OCT-2001]	153..296 2..148	76/148 (51%) 100/148 (67%)	8e-34
AAU58306	Propionibacterium acnes immunogenic protein #19202 - Propionibacterium acnes, 246 aa. [WO200181581-A2, 01-NOV-2001]	154..289 79..218	69/142 (48%) 94/142 (65%)	8e-30
ABG03944	Novel human diagnostic protein #3935 - Homo sapiens, 135 aa. [WO200175067-A2, 11-OCT-2001]	95..221 5..131	72/130 (55%) 85/130 (65%)	1e-29

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7D.

Table 7D. Public BLASTP Results for NOV7a

Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UH96	LRP16 (LRP16 protein) - Homo sapiens (Human), 325 aa.	1..313 1..325	313/325 (96%) 313/325 (96%)	e-179
Q9BQ69	LRP16 protein - Homo sapiens (Human), 243 aa.	83..313 1..243	231/243 (95%) 231/243 (95%)	e-130
Q922B1	Similar to LRP16 protein - Mus musculus (Mouse), 243 aa.	83..313 1..243	210/243 (86%) 221/243 (90%)	e-118
AAM45760	LRP16-like protein - Rattus norvegicus (Rat), 243 aa.	83..313 1..243	206/243 (84%) 218/243 (88%)	e-115
AAM73435	Histone macro-H2A1-related protein - Chlorobium tepidum TLS, 172 aa.	151..306 4..169	91/167 (54%) 114/167 (67%)	6e-43

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7E.

Table 7E. Domain Analysis of NOV7a			
Pfam Domain	**NOV7a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
A1pp	170..284	53/124 (43%) 94/124 (76%)	1.3e-45

Example 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 27	732 bp	
NOV8a, CG145988-01 DNA Sequence	TGACATAAGATACCATGAGTAAAATACTGAGAACACAGTAAGCCCTCTGTAATTTTTCAGCTGGCGGA GACCCCCGTGTCGCCGCTGCCGCCGGCGCCTGGCCTCCCGCGGCACTCCCGGCTGCGCGCTCCGCTCG GCCCCGCCATGCGGGTGAACCTCAGACAGCGGTCCGGGCCCTGGAGCCTTCAGGAAGTGGACGAACAG CCGCAGCACCCGCTGCATGTCACCTGCACGGTGGCGGTGGACGAGCGGAGCACCGTGCCCGATCCCAC CCAGGTGAAGAACAGACCCACCAGCATTTCGTGGGATGGTCTTGATTTAGGGAAACTCTACACCTGGG TCTTGACAGACCCGGATGCTCCAAGCAGGAAGGATCCCCAATACAGGGAATGGCATCATTTCCTGGTG GTCCACATGAAGGGCAGTGACATCAGCAGTGGCACAGTCCTCTCCGGATTATGTGGGCTCGGGGTCTC CCAAGGCACAGGCATGCACCACCACGCCTGGCTAATTTTTCCAAGGCACAGGCCGCTGAAATGTGACG AGCCCATCCTCAGCAACCGCTCCGGAGACCACCGTGGCAAATTCAAGGTGGCGTCCCCCCGTAAAAAG TACGCGCTCGGGGTCCCGGTGGCCGGCACGTGTTACCAGGCCGAGTGGGACGACTATGTGCCCAAATT GTAGGAGCAGCTGTCTGGGAAGTAGGGGGCAGCTCGGCGACCTGAACCGTCC		
	ORF Start: ATG at 145		ORF Stop: TAG at 682
	SEQ ID NO: 28	179 aa	MW at 20196.8kD
NOV8a, CG145988-01 Protein Sequence	MRVNLRQRSGPWSLQEVDEQPQHPLHVTCTVAVDERSTVPDPTQVKNRPTSISWDGLDLGKLYTWVLT DPDAPSRKDPQYREWHHFLVVHMKGSDISSGTVLSGLCGLGVSQGTGMHHHAWLIFPRHRPLKCDEPI LSNRSGDHRGKFKVASPRKKYALGVPVAGTCYQAEWDDYVPKL		

Further analysis of the NOV8a protein yielded the following properties shown in Table 8B.

Table 8B. Protein Sequence Properties NOV8a	
PSort analysis:	0.5277 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1725 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV8a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8C.

Table 8C. GENESEQ Results for NOV8a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAR64268	Phosphatidylethanolamine binding protein - Homo sapiens, 187 aa. [EP628631-A, 14-DEC-1994]	1..179 1..180	141/180 (78%) 152/180 (84%)	9e-79
AAE21677	Human phosphoethanolamine binding protein (PEBP) - Homo sapiens, 187 aa. [WO200218623-A2, 07-MAR-2002]	1..179 1..180	140/180 (77%) 151/180 (83%)	5e-78
AAR49943	Human hippocampal cholinergic neurotrophic peptide precursor - Homo sapiens, 187 aa. [WO9405788-A, 17-MAR-1994]	1..179 1..180	140/180 (77%) 151/180 (83%)	5e-78
AAR27718	HCNP precursor protein #2 - Homo sapiens, 187 aa. [EP511816-A, 04-NOV-1992]	1..179 1..180	140/180 (77%) 151/180 (83%)	5e-78
AAE21676	Mouse phosphoethanolamine binding protein (PEBP) - Mus musculus, 187 aa. [WO200218623-A2, 07-MAR-2002]	1..179 1..180	125/180 (69%) 142/180 (78%)	1e-69

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8D.

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	**Protein/Organism/Length**	**NOV8a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
AAH31102	Prostatic binding protein - Homo sapiens (Human), 187 aa.	1..179 1..180	140/180 (77%) 151/180 (83%)	1e-77
S46485	phosphatidylethanolamine-binding protein - crab-eating macaque, 187 aa.	1..179 1..180	139/180 (77%) 151/180 (83%)	2e-77
P30086	Phosphatidylethanolamine-binding protein (PEBP) (Neuropolypeptide h3) (Hippocampal cholinergic neurostimulating peptide) (HCNP) (Raf kinase inhibitor protein) (RKIP) - Homo sapiens (Human), 186 aa.	3..179 2..179	139/178 (78%) 150/178 (84%)	3e-77

P48737	Phosphatidylethanolamine-binding protein (PEBP) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 186 aa.	3..179 2..179	138/178 (77%) 150/178 (83%)	4e-77
P13696	Phosphatidylethanolamine-binding protein (PEBP) (Basic cytosolic 21 kDa protein) - Bos taurus (Bovine), 186 aa.	3..179 2..179	134/178 (75%) 148/178 (82%)	1e-73

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8E.

Table 8E. Domain Analysis of NOV8a

Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PBP	3..171	71/201 (35%) 137/201 (68%)	4.7e-41

Example 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis

	SEQ ID NO: 29	704 bp	
NOV9a, CG146452-01 DNA Sequence	CTCTTCCACCGGCCCCTCAAGGAGTACTCCTTCCGTTCCGTGCGGGAGGAGACTGGCATGGGGGACAT TCCTGACGTCAAGAATGACTTCGCCTTCATGCTGCACCTCATCGATCAGTACGACTCCCTCTACTCCA AGCGCTTCGCCGTCTTCCTGTCCGAGGTCAGCGAAAGCCGTCTAAAGCAGCTCTACCTCAGCTACAAC AAGCTGGAGACCCTGCCCTCCCAGCTCGGCCTGTGCTCAGGCCTCCGTCTGCTGGATGTGTCCCACAA TGGGCTACACTCCCTGCCACCCGAGGTGGGCCTCCTGCAGAACCTACAGCACCTGGCCCTCTCCTACA ATGCCCTGGAGGCCCTGCCCGAAGAGCTCTTCTTCTGCCGCAAGCTGCGGACGTTGCTTCTGGGCGAC AACCAGCTGAGCCAGCTCTCGCCCCACGTGGGTGCCCTCAGAGCCCTCAGCCGCCTGGAGCTCAAAGG CAACCGCTTAGAGGCGCTGCCAGAAGAACTTGGCAACTGTGGGGGGCTCAAGAAGGCGGGGCTCCTGG TGGAAGACACGCTTTACCAGGGTCTGCCGGCAGAAGTGCGGGACAAGATGGAGGAGGAATGAAGCTGG GGTGGGGCCGTTTTAGGTAGAGCCTTAAAAATGCTTCTGTCCTGGAATCTCAACCATTGTCTTCCAAG ATAGGAAGCCAAGTGGGTCTAGGC		
	ORF Start: ATG at 58		ORF Stop: TGA at 604
	SEQ ID NO: 30	182 aa	MW at 20391.2kD
NOV9a, CG146452-01 Protein Sequence	MGDIPDVKNDFAFMLHLIDQYDSLYSKRFAVFLSEVSESRLKQLYLSYNKLETLPSQLGLCSGLRLLD VSHNGLHSLPPEVGLLQNLQHLALSYNALEALPEELFFCRKLRTLLLGDNQLSQLSPHVGALRALSRL ELKGNRLEALPEELGNCGGLKKAGLLVEDTLYQGLPAEVRDKMEEE		

Further analysis of the NOV9a protein yielded the following properties shown in Table 9B.

Table 9B. Protein Sequence Properties NOV9a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9C.

Table 9C. GENESEQ Results for NOV9a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV9a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAM39906	Human polypeptide SEQ ID NO 3051 - Homo sapiens, 543 aa. [WO200153312-A1, 26-JUL-2001]	41..182 402..543	74/142 (52%) 111/142 (78%)	2e-39
AAU17987	Human immunoglobulin polypeptide SEQ ID No 132 - Homo sapiens, 277 aa. [WO200155315-A2, 02-AUG-2001]	41..182 136..277	74/142 (52%) 111/142 (78%)	2e-39
AAY70473	Human cyclic nucleotide-associated protein-1 (CNAP-1) - Homo sapiens, 708 aa. [WO200014248-A1, 16-MAR-2000]	41..182 567..708	74/142 (52%) 111/142 (78%)	2e-39
AAM41692	Human polypeptide SEQ ID NO 6623 - Homo sapiens, 565 aa. [WO200153312-A1, 26-JUL-2001]	41..182 424..565	74/142 (52%) 110/142 (77%)	6e-39
AAU83653	Human PRO protein, Seq ID No 124 - Homo sapiens, 546 aa. [WO200208288-A2, 31-JAN-2002]	39..179 398..538	76/141 (53%) 104/141 (72%)	3e-37

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9D.

Table 9D. Public BLASTP Results for NOV9a				
Protein Accession Number	**Protein/Organism/Length**	**NOV9a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**

CAD39133	Hypothetical protein - Homo sapiens (Human), 622 aa (fragment).	41..182 481..622	141/142 (99%) 142/142 (99%)	1e-75
Q9H5H8	CDNA: FLJ23420 fis, clone HEP22352 (Hypothetical 50.2 kDa protein) - Homo sapiens (Human), 444 aa.	41..182 303..444	141/142 (99%) 142/142 (99%)	1e-75
AAH31863	Hypothetical protein - Mus musculus (Mouse), 330 aa (fragment).	41..182 189..330	118/142 (83%) 132/142 (92%)	2e-62
Q8R502	AD158 - Mus musculus (Mouse), 803 aa.	41..182 662..803	73/142 (51%) 112/142 (78%)	2e-39
Q8VE36	Similar to hypothetical protein DKFZp586J1119 - Mus musculus (Mouse), 287 aa (fragment).	41..182 146..287	73/142 (51%) 112/142 (78%)	2e-39

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9E.

Table 9E. Domain Analysis of NOV9a

Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR	40..62	9/25 (36%) 21/25 (84%)	0.0068
LRR	63..85	13/25 (52%) 17/25 (68%)	0.15
LRR	86..108	10/25 (40%) 18/25 (72%)	0.33
LRR	109..131	10/25 (40%) 19/25 (76%)	0.62

Example 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis

	SEQ ID NO: 31	2235 bp
NOV10a, CG146731-01 DNA Sequence	GTTTCCTGTTATACTGCTCTCCAATCCATCATGAACCAGCCAGAGTCTGCCAACGATCCTGAACCCCT GTGTGCAGTGTGTGGCCAAGCCCACTCCTTGGAGGAAAACCACTTCTACAGCTATCCAGAGGAAGTGG ATGATGACCTCATCTGCCACATCTGCCTGCAGGCTTTGCTGGACCCCCTGGACACTCCGTGTGGACAC ACCTACTGCACCCTCTGCCTCACCAACTTCCTGGTGGAGAAGGACTTCTGTCCCATGGACCGCAAGCC TCTGGTTCTGCAGCACTGCAAGAAGTCCAGCATCCTGGTCAACAAACTCCTCAACAAGCTACTGGTGA CCTGCCCATTCAGGGAGCACTGCACCCAGGTGTTGCAGCGCTGTGACCTCGAGCATCACTTTCAAACC AGCTGTAAAGGTGCCTCCCACTACGGCCTGACCAAAGATAGGAAGAGGCGCTCACAAGATGGCTGTCC AGACGGCTGTGCGAGCCTCACAGCCACGGCTCCCTCCCCAGAGGTTTCTGCAGCTGCCACCATCTCCT TAATGACAGACGAGCCTGGCCTAGACAACCCTGCCTACGTGTCCTCGGCAGAGGACGGGCAGCCAGCA	

	ATCAGCCCAGTGGACTCTGGCCGGAGCAACCGAACTAGGGCACGGCCCTTTGAGAGATCCACTATTAG AAGCAGATCATTTAAAAAAATAAATCGAGCTTTGAGTGTTCTTCGAAGGACAAAGAGCGGGAGTGCAG TTGCCAACCATGCCGACCAGGGCAGGGAAAATTCTGAAAACACCACTGCCCCTGAAGTCTTTCCAAGG TTGTACCACCTGATTCCAGATGGTGAAATTACCAGCATCAAGATCAATCGAGTAGATCCCAGTGAAAG CCTCTCTATTAGGCTGGTGGGAGGTAGCGAAACCCCACTGGTCCATATCATTATCCAACACATTTATC GTGATGGGGTGATCGCCAGAGACGGCCGGCTACTGCCAGGAGACATCATTCTAAAGGTAAACGGGATG GACATCAGCAATGTCCCTCACAACTACGCTGTGCGCTCTCCTGCGGCACCCTGCCAGGTGCTGTGGCT GACTGTGATGCGTGAACAGAAGTTCCGCAGCAGGAACAATGGACAGGCCCCGGATGCCTACAGACCCC GAGATGACAGCTTTCATGTGATTCTCAACAAAAGTAGCCCCGAGGAGCAGCTTGGAATAAAACTGGTG CGCAAGGTGGATGAGCCTGGGGTTTTCATCTTCAATGTGCTGGATGGCGGTGTGGCATATCGACATGG TCAGCTTGAGGAGAATGACCGTGTGTTAGCCATCAATGGACATGATCTTCGATATGGCAGCCCAGAAA GTGCGGCTCATCTGATTCAGGCCAGTGAAAGACGTGTTCACCTCGTCGTGTCCCGCCAGGTTCGGCAG CGGAGCCCTGACATCTTTCAGGAAGCCGGCTGGAACAGCAATGGCAGCTGGTCCCCAGGGCCAGGGGA GAGGAGCAACACTCCCAAGCCTACAATTACTTGTCATGAGAAGGTGGTAAATATCCAAAAAGACCCCG GTGAATCTCTCGGCATGACCGTCGCAGGGGGAGCATCACATAGAGAATGGGATTTGCCTATCTATGTC ATCAGTGTTGAGCCCGGAGGAGTCATAAGCAGAGATGGAAGAATAAAAACAGGTGACATTTTGTTGAA TGTGGATGGGGTCGAACTGACAGAGGTCAGCCGGAGTGAGGCAGTGGCATTATTGAAAAGAACATCAT CCTCGATAGTACTCAAAGCTTTGGAAGTCAAAGAGTATGAGCCCCAGGAAGACTGCAGCAGCCCAGCA GCCCTGGACTCCAACCACAACATGGCCCCACCCAGTGACTGGTCCCCATCCTGGGTCATGTGGCTGGA ATTACCACGGTGCTTGTATAACTGTAAAGATATTGTATTACGAAGAAACACAGCTGGAAGTCTGGGCT TCTGCATTGTAGGAGGTTATGAAGAATACAATGGAAACAAACCTTTTTTCATCAAATCCATTGTTGAA GGAACACCAGCATACAATGATGGAAGAATTAGGTGTGGTGATATTCTTCTTGCTGTCAATGGTAGAAG TACATCAGGAATGATACATGCTTGCTTGGCAAGACTGCTGAAAGAACTTAAAGGAAGAATTACTCTAA CTATTGTTTCTTGGCCTGGCACTTTTTTATAGAATCAATGATGGGTCAGAGGAAAACAG
	ORF Start: ATG at 31		ORF Stop: TAG at 2206
	SEQ ID NO: 32	725 aa	MW at 80095.8kD
NOV10a, CG146731-01 Protein Sequence	MNQPESANDPEPLCAVCGQAHSLEENHFYSYPEEVDDDLICHICLQALLDPLDTPCGHTYCTLCLTNF LVEKDFCPMDRKPLVLQHCKKSSILVNKLLNKLLVTCPFREHCTQVLQRCDLEHHFQTSCKGASHYGL TKDRKRRSQDGCPDGCASLTATAPSPEVSAAATISLMTDEPGLDNPAYVSSAEDGQPAISPVDSGRSN RTRARPFERSTIRSRSFKKINRALSVLRRTKSGSAVANHADQGRENSENTTAPEVFPRLYHLIPDGEI TSIKINRVDPSESLSIRLVGGSETPLVHIIIQHIYRDGVIARDGRLLPGDIILKVNGMDISNVPHNYA VRSPAAPCQVLWLTVMREQKFRSRNNGQAPDAYRPRDDSFHVILNKSSPEEQLGIKLVRKVDEPGVFI FNVLDGGVAYRHGQLEENDRVLAINGHDLRYGSPESAAHLIQASERRVHLVVSRQVRQRSPDIFQEAG WNSNGSWSPGPGERSNTPKPTITCHEKVVNIQKDPGESLGMTVAGGASHREWDLPIYVISVEPGGVIS RDGRIKTGDILLNVDGVELTEVSRSEAVALLKRTSSSIVLKALEVKEYEPQEDCSSPAALDSNHNMAP PSDWSPSWVMWLELPRCLYNCKDIVLRRNTAGSLGFCIVGGYEEYNGNKPFFIKSIVEGTPAYNDGRI RCGDILLAVNGRSTSGMIHACLARLLKELKGRITLTIVSWPGTFL	

Further analysis of the NOV10a protein yielded the following properties shown in Table 10B.

Table 10B. Protein Sequence Properties NOV10a

PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV10a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10C.

Table 10C. GENESEQ Results for NOV10a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAB65212	Human PRO1136 (UNQ574) protein sequence SEQ ID NO:219 - Homo sapiens, 632 aa. [WO200073454-A1, 07-DEC-2000]	101..725 5..632	598/628 (95%) 603/628 (95%)	0.0
AAB87545	Human PRO1136 - Homo sapiens, 632 aa. [WO200116318-A2, 08-MAR-2001]	101..725 5..632	598/628 (95%) 603/628 (95%)	0.0
AAU17394	Novel signal transduction pathway protein, Seq ID 959 - Homo sapiens, 696 aa. [WO200154733-A1, 02-AUG-2001]	101..725 69..696	598/628 (95%) 603/628 (95%)	0.0
AAU29106	Human PRO polypeptide sequence #83 - Homo sapiens, 632 aa. [WO200168848-A2, 20-SEP-2001]	101..725 5..632	598/628 (95%) 603/628 (95%)	0.0
AAY66689	Membrane-bound protein PRO1136 - Homo sapiens, 632 aa. [WO9963088-A2, 09-DEC-1999]	101..725 5..632	598/628 (95%) 603/628 (95%)	0.0

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10D.

Table 10D. Public BLASTP Results for NOV10a				
Protein Accession Number	**Protein/Organism/Length**	**NOV10a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q8TBB1	Similar to multi-PDZ-domain-containing protein - Homo sapiens (Human), 728 aa.	1..725 1..728	721/728 (99%) 721/728 (99%)	0.0
Q96MJ7	CDNA FLJ32261 fis, clone PROST1000343, highly similar to Numb-binding protein LNXp80 - Homo sapiens (Human), 728 aa.	1..725 1..728	718/728 (98%) 718/728 (98%)	0.0
O70263	NUMB-binding protein LNX (Ligand of NUMB-protein X) - Mus musculus (Mouse), 728 aa.	1..725 1..728	640/729 (87%) 673/729 (91%)	0.0
AAH34737	Ligand of numb-protein X - Homo sapiens (Human), 632 aa.	101..725 5..632	598/628 (95%) 603/628 (95%)	0.0

Q9BY20	Multi-PDZ-domain-containing protein - Homo sapiens (Human), 632 aa.	101..725 5..632	597/628 (95%) 602/628 (95%)	0.0

PFam analysis predicts that the NOV10a protein contains the domains shown in the Table 10E.

Table 10E. Domain Analysis of NOV10a

Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C3HC4	41..77	14/53 (26%) 26/53 (49%)	9.8e-06
PDZ	274..358	26/86 (30%) 62/86 (72%)	5.7e-11
PDZ	381..463	28/84 (33%) 57/84 (68%)	4.7e-12
PDZ	504..589	25/88 (28%) 63/88 (72%)	3.8e-13
PDZ	635..720	26/88 (30%) 66/88 (75%)	1.3e-14

Example 11.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis

	SEQ ID NO: 33	1344 bp	
NOV11a, CG147048-01 DNA Sequence	ATGGACTCAGACTTCTCACATGCCTTCCAGAAGGAACTCACCTGTGTCATCTGTTTGAACTACCTGGT AGACCCTGTCACCATCTGCTGTGGGCACAGCTTCTGTAGGCCCTGTCTCTGCCTTTCGTGGGAGGAAG CCCAAAGTCCTGCAAACTGCCCTGCATGCAGGGAACCATCACCGAAAATGGACTTCAAAACCAATATT CTTCTGAAGAATTTAGTGACCATTGCCAGAAAAGCCAGTCTCTGGCAATTCCTGAGCTCTGAGAAACA AATATGTGGGACCCATAGGCAAACAAAGAAGATGTTCTGTGACATGGACAAGAGTCTCCTCTGCTTGC TGTGCTCCAACTCTCAGGAGCACGGGGCTCACAAACACCATCCCATCGAAGAGGCAGCTGAGGAACAC CGGGAGAAACTCTTAAAGCAAATGAGGATTTTATGGAAAAAGATTCAAGAAAATCAGAGAAATCTATA TGAGGAGGGAAGAACAGCCTTCCTCTGGAGGGGCAATGTGGTTTTACGGGCACAGATGATCAGGAATG AGTATAGGAAGCTGCATCCGGTTCTCCATAAGGAAGAAAAACAACATTTAGAGAGACTGAACAAGGAA TACCAAGAGATTTTTCAGCAACTCCAGAGAAGTTGGGTCAAAATGGATCAAAAGAGTAAACACTTGAA AGAAATGTATCAGGAACTAATGGAAATGTGTCATAAACCAGATGTGGAGCTGCTCCAGAGTGAGTCCG TGCTGCTGCACATGCCCCAGCCTGTGAATCCAGAGCTCACTGCAGGACCCATCACTGGACTGGTGTAC AGGCTCAACCGCTTCCGAGTGGAAATTTCCTTCCATTTTGAAGTAACCAATCACAATATCAGGCTCTT TGAGGATGTGAGAAGTTGGATGTTTAGACGTGGACCTTTGAATTCTGACAGATCTGACTATTTTGCTG CATGGGGAGCCAGGGTCTTCTCCTTTGGGAAACACTACTGGGAGCTGGATGTGGACAACTCTTGTGAC TGGGCTCTGGGAGTCTGTAACAACTCCTGGATAAGGAAGAATAGCACAATGGTTAACTCTGAGGACAT ATTTCTTCTTTTGTGTCTGAAGGTGGATAATCATTTCAATCTCTTGACCACCTCCCCAGTGTTTCCTC ACTACATAGAGAAACCTCTGGGCCGGGTTGGTGTGTTTCTTGATTTTGAAAGTGGAAGTGTGAGTTTT TTGAATGTCACCAAGAGTTCCCTCATATGGAGTTACCCAGCTGGCTCCTTAACTTTTCCTGTCAGGCC TTTCTTTTACACTGGCCACAGATGATCAGGATTAAGAAAACTTACTGTTTGG		
	ORF Start: ATG at 1		ORF Stop: TGA at 1315

	SEQ ID NO: 34	438 aa	MW at 51424.8kD
NOV11a, CG147048-01 Protein Sequence	MDSDFSHAFQKELTCVICLNYLVDPVTICCGHSFCRPCLCLSWEEAQSPANCPACREPSPKMDFKTNI LLKNLVTIARKASLWQFLSSEKQICGTHRQTKKMFCDMDKSLLCLLCSNSQEHGAHKHHPIEEAAEEH REKLLKQMRILWKKIQENQRNLYEEGRTAFLWRGNVVLRAQMIRNEYRKLHPVLHKEEKQHLERLNKE YQEIFQQLQRSWVKMDQKSKHLKEMYQELMEMCHKPDVELLQSESVLLHMPQPVNPELTAGPITGLVY RLNRFRVEISFHFEVTNHNIRLFEDVRSWMFRRGPLNSDRSDYFAAWGARVFSFGKHYWELDVDNSCD WALGVCNNSWIRKNSTMVNSEDIFLLLCLKVDNHFNLLTTSPVFPHYIEKPLGRVGVFLDFESGSVSF LNVTKSSLIWSYPAGSLTFPVRPFFYTGHR		

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3233 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV11a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11C.

Table 11C. GENESEQ Results for NOV11a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG15486	Novel human diagnostic protein #15477 - Homo sapiens, 414 aa. [WO200175067-A2, 11-OCT-2001]	56..246 224..414	191/191 (100%) 191/191 (100%)	e-111
ABG15487	Novel human diagnostic protein #15478 - Homo sapiens, 369 aa. [WO200175067-A2, 11-OCT-2001]	24..169 54..199	139/146 (95%) 142/146 (97%)	2e-81
AAR15148	Ro/SSA autoantigen - Homo sapiens, 475 aa. [WO9117171-A, 14-NOV-1991]	11..436 12..449	146/443 (32%) 226/443 (50%)	4e-54
AAB42919	Human ORFX ORF2683 polypeptide sequence SEQ ID NO:5366 - Homo sapiens, 477 aa. [WO200058473-A2, 05-OCT-2000]	4..436 5..459	138/461 (29%) 230/461 (48%)	4e-53
AAM48396	Human SSA-56kDa protein - Homo sapiens, 485 aa. [WO200188128-A1, 22-NOV-2001]	8..432 9..454	135/454 (29%) 217/454 (47%)	5e-47

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11D.

Table 11D. Public BLASTP Results for NOV11a

Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96BQ3	Hypothetical 52.3 kDa protein - Homo sapiens (Human), 446 aa.	1..438 1..446	438/446 (98%) 438/446 (98%)	0.0
Q9NS80	RING finger protein 18 (Testis-specific ring-finger protein) - Homo sapiens (Human), 452 aa.	1..437 1..451	230/451 (50%) 302/451 (65%)	e-125
BAC04185	CDNA FLJ36180 fis, clone TESTI2026605, weakly similar to 52 KDA RO PROTEIN - Homo sapiens (Human), 468 aa.	3..433 4..445	148/445 (33%) 236/445 (52%)	7e-64
Q9BSJ1	Similar to ring finger protein 18 - Homo sapiens (Human), 293 aa.	168..437 9..292	128/284 (45%) 182/284 (64%)	3e-63
AAM63957	BIA1 protein - Homo sapiens (Human), 468 aa.	4..433 5..442	150/446 (33%) 231/446 (51%)	3e-62

PFam analysis predicts that the NOV11a protein contains the domains shown in the Table 11E.

Table 11E. Domain Analysis of NOV11a

Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C3HC4	15..55	18/54 (33%) 35/54 (65%)	5.7e-10
zf-B_box	88..129	14/48 (29%) 26/48 (54%)	1.3e-06
SPRY	326..437	28/157 (18%) 77/157 (49%)	5.8e-09

Example 12.

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis

	SEQ ID NO: 35	2711 bp	
NOV12a, CG147246-01 DNA Sequence	ATGGAGGAAATTAAACCTGCCTCTGCTTCTTGTGTCTCAAAAGAAAAACCCAGTAAGGTATCAGATCT CATCAGTCGCTTTGAAGGAGGCAGCTCATTATCAAATTATAGTGATTTGAAGAAAGAGTCTGCTGTGA ACCTAAATGCTCCTAGAACCCCAGGAAGGCATGGATTGACAACCACACCTCAACAAAAACTCCTCTCC CAGCACTTGCCACAGAGGCAGGGAAATGATACAGATAAGACTCAGGGTGCACAGACTTGTGTGGCCAA CGGTGTAATGGCAGCACAAAACCAGATGGAATGTGAGGAGGAGAAAGCTGCCACTCTTAGCTCAGATA CTTCTATTCAAGCTTCTGAACCCTTGCTTGATACGCACATAGTGAATGGAGAAAGAGATGAAACTGCC ACAGCTCCTGCATCACCCACAACAGACAGCTGTGATGGAAATGCTTCTGACAGTAGCTACAGGACTCC AGGCATAGGCCCAGTGCTCCCCCTAGAAGAAAGAGGGGCAGAAACAGAAACCAAGGTACAAGAGAGGG AAAATGGGGAAAGCCCTCTGGAACTGGAGCAGCTGGACCAGCACCATGAGATGAAGGAGACTAATGAG CAAAAACTTCACAAAATAGCCAATGAACTTTGCTTACTGAAAGAGCTTATGTCAACCGACTTGACCT CTTAGATCAGGTAGTATTTTATTGCAAACTGTTGGAAGAAGCAAACCGAGGCTCGTTTCCAGCAGAGA TGGTGAATAAAATCTTTTCTAATATTTCATCAATAAATGCCTTCCATAGTAAATTCCTCTTGCCAGAG CTGGAGAAACGAATGCAAGAATGGGAAACTACTCCTAGAATTGGAGACATCCTTCAGAAATTGGCACC ATTCCTTAAGATGTATGGAGAATATGTGAAAGGATTTGATAATGCAATGGAATTGGTTAAAAACATGA CAGAACGTATTCCCCAGTTCAAATCAGTGGTTGAAGAAATTCAGAAACAGAAAATCTGTGGGAGCTTA ACTTTGCAGCATCACATGCTAGAACCTGTTCAGCGGATTCCCCGGTATGAGATGCTCCTTAAGGACTA TCTAAGGAAATTGCCTCCTGATTCCCTGGACTGGAATGATGCTAAAGAATCACTTGAAATTATATCTA CAGCAGCAAGCCATTCTAATAGTGCAATAAGGAAAATGGAGAACCTAAAGAAACTCTTAGAGATTTAT GAAATGTTGGGAGAAGAAGAAGACATTGTGAACCCTTCAAATGAACTAATAAATGAAGGACAGATCCT CAAACTAGCTGCTCGGAACACTTCAGCACAAGAACGCTACCTTTTCTTATTCAACAACATGTTGCTGT ACTGTGTGCCCAAATCCAGCTTGGTAGGCTCTAAATTCACAGTTCGAACCAGGGTTGGCATTGATGGA ATGAAAATTGTAGAGACTCAAAATGAAGAATATCCACATACTTTCCAGGTGTCTGGGAAAGAGAGAAC ACTGGAACTGCAGGCCAGTTCTGCGCAAGACAAAGAAGAATGGATCAAGGCCCTTCAAGAAACCATCG ATGCTTTTCATCAAAGGCATGAAACCTTCAGAAATGCAATTGCAAAGGATAATGACATTCACTCAGAG GTTTCTACTGCTGAGCTAGGGAAAAGAGCCCCAAGATGGATCCGAGATAATGAAGTGACAATGTGTAT GAAATGTAAAGAACCTTTCAATGCACTGACACGAAGGAGGCATCATTGTCGAGCATGTGGATATGTGG TTTGTTGGAAATGCTCCGACTACAAAGCTCAACTTGAATATGATGGTGGTAAATTGAGCAAAGTTTGT AAAGACTGTTATCAAATCATAAGTGGATTCACAGACAGTGAAGAAAAGAAAAGAAAAGGAATTTTAGA GATTGAATCAGCAGAAGTATCTGGAAACAGTGTGGTGTGCAGCTTTCTTCAGTATATGGAGAAGTCAA AACCTTGGCAGAAAGCTTGGTGTGTGATCCCCAAGCAAGACCCTCTTGTGCTGTACATGTATGGTGCC CCCCAGGACGTCAGAGCCCAGGCCACCATTCCACTTCTGGGCTATGTGGTGGATGAAATGCCAAGGAG CGCAGACCTGCCACACAGTTTCAAACTGACCCAGTCTAAGTCCGTGCACAGCTTTGCTGCAGACAGTG AGGAACTGAAGCAGAAGTGGCTGAAAGTCATCCTTTTAGCTGTCACAGGTGAGACACCAGGTGGTCCA AATGAGCATCCAGCCACCTTGGATGATCATCCTGAACCTAAGAAAAAATCAGAATGCTGAACTCCTCC AGGACCAGCCATGGTGTGGAGGTCTCAGGACTTACAGCTCAAGACATTCCCAGCTCTTCTTACACATC TGCTAGCACTTTATGTTGAAAAATATAGGCCCATAAATGCATCTTTTGAGGACTATTTTCCTATGTTT ATGTACTCTTAGTGAAATTAGTGTGCAGAGTCATTCTACCGATAAAGTTTTGAAATAATGTGAAAACT GGAGCATTTTTTGAGCTATTCCTTGAATATGTGCTTTTTTGTCTTGAAGAAATGGTGTATCAATTGAT TCTGTCACCGTCAGGTTAGAATGAGCACTTCCATTTAAGAAATCCTTTCATGTCTTCTTCTCTTTCAC ATGTAGGACCTGGAACAGTTTGAAAGATATACCTCCATGTTGCCAAAATAGATCCATGG		
	ORF Start: ATG at 1		ORF Stop: TGA at 2302
	SEQ ID NO: 36	767 aa	MW at 86651.3kD
NOV12a, CG147246-01 Protein Sequence	MEEIKPASASCVSKEKPSKVSDLISRFEGGSSLSNYSDLKKESAVNLNAPRTPGRHGLTTTPQQKLLS QHLPQRQGNDTDKTQGAQTCVANGVMAAQNQMECEEEKAATLSSDTSIQASEPLLDTHIVNGERDETA TAPASPTTDSCDGNASDSSYRTPGIGPVLPLEERGAETETKVQERENGESPLELEQLDQHHEMKETNE QKLHKIANELLLTERAYVNRLDLLDQVVFYCKLLEEANRGSFPAEMVNKIFSNISSINAFHSKFLLPE LEKRMQEWETTPRIGDILQKLAPFLKMYGEYVKGFDNAMELVKNMTERIPQFKSVVEEIQKQKICGSL TLQHHMLEPVQRIPRYEMLLKDYLRKLPPDSLDWNDAKESLEIISTAASHSNSAIRKMENLKKLLEIY EMLGEEEDIVNPSNELINEGQILKLAARNTSAQERYLFLFNNMLLYCVPKSSLVGSKFTVRTRVGIDG MKIVETQNEEYPHTFQVSGKERTLELQASSAQDKEEWIKALQETIDAFHQRHETFRNAIAKDNDIHSE VSTAELGKRAPRWIRDNEVTMCMKCKEPFNALTRRRHHCRACGYVVCWKCSDYKAQLEYDGGKLSKVC KDCYQIISGFTDSEEKKRKGILEIESAEVSGNSVVCSFLQYMEKSKPWQKAWCVIPKQDPLVLYMYGA PQDVRAQATIPLLGYVVDEMPRSADLPHSFKLTQSKSVHSFAADSEELKQKWLKVILLAVTGETPGGP NEHPATLDDHPEPKKKSEC		

Further analysis of the NOV12a protein yielded the following properties shown in Table 12B.

Table 12B. Protein Sequence Properties NOV12a	
PSort analysis:	0.7000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV12a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12C.

Table 12C. GENESEQ Results for NOV12a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU79515	Human guanine-nucleotide exchange factor (GEF), 38646 - Homo sapiens, 766 aa. [WO200220765-A2, 14-MAR-2002]	1..767 1..766	763/767 (99%) 764/767 (99%)	0.0
AAY51248	Rat actin-binding protein frabin - Rattus sp, 766 aa. [JP11346775-A, 21-DEC-1999]	1..767 1..766	634/767 (82%) 682/767 (88%)	0.0
AAU21630	Novel human neoplastic disease associated polypeptide #63 - Homo sapiens, 465 aa. [WO200155163-A1, 02-AUG-2001]	303..767 1..465	459/465 (98%) 460/465 (98%)	0.0
AAU17094	Novel signal transduction pathway protein, Seq ID 659 - Homo sapiens, 319 aa. [WO200154733-A1, 02-AUG-2001]	467..767 19..319	298/301 (99%) 298/301 (99%)	e-178
AAU27818	Human full-length polypeptide sequence #143 - Homo sapiens, 725 aa. [WO200164834-A2, 07-SEP-2001]	164..744 120..706	278/609 (45%) 378/609 (61%)	e-142

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12D.

Table 12D. Public BLASTP Results for NOV12a				
Protein Accession Number	**Protein/Organism/Length**	**NOV12a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q96M96	CDNA FLJ32732 fis, clone TESTI2001141, highly similar to Rattus norvegicus actin-filament binding protein Frabin mRNA - Homo sapiens (Human), 766 aa.	1..767 1..766	762/767 (99%) 763/767 (99%)	0.0
Q91ZT5	Actin-binding protein frabin-alpha - Mus musculus (Mouse), 766 aa.	1..767 1..766	642/767 (83%) 692/767 (89%)	0.0
O88387	Actin-filament binding protein Frabin - Rattus norvegicus (Rat), 766 aa.	1..767 1..766	634/767 (82%) 682/767 (88%)	0.0
Q91ZT4	Actin-binding protein frabin-beta - Mus musculus (Mouse), 603 aa.	1..589 1..588	480/589 (81%) 520/589 (87%)	0.0
Q91ZT3	Actin-binding protein frabin-gamma - Mus musculus (Mouse), 504 aa.	1..504 1..503	402/504 (79%) 439/504 (86%)	0.0

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12E.

Table 12E. Domain Analysis of NOV12a			
Pfam Domain	**NOV12a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
RhoGEF	210..393	75/209 (36%) 155/209 (74%)	4.3e-57
PH	424..522	30/99 (30%) 80/99 (81%)	1.5e-16
DAG_PE-bind	556..602	12/51 (24%) 28/51 (55%)	0.63
FYVE	555..621	32/72 (44%) 54/72 (75%)	1.1e-23
PH	645..741	23/97 (24%) 69/97 (71%)	1.2e-12

Example 13.

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 37	1218 bp	
NOV13a, CG147651-01 DNA Sequence	GAGCCCGTCCGGACTTCCCCGATCCCAGCCTTCTCTCCTTTGAAAACACTAAGAATGACATCACTGCA TCAGTTTTTACTAGAGCCAATCACCTGTCGTGCCTGGAACAGGGATCGTACCCAGATTGCCCTCAGTC CTGATAATCAGGAAGCGCACATCTATAAGAAGCATGGGAGCCAGCGAGTGAAAGCTCGTGAACTCAGT GAGCACAATGGACATATCACAGGCATTCACTGGGCTCCCAAGAGCGACCGCATCGTCACTTGCGGGGT AGACTGCAATGCCTATGTCTGGAGTCAGAAAGATGGTGTCTGGAAGCCAACCCTGGTGATCCTTAGAA TTAATCGTGCAGCTACTTTTGTGAAGTGGTCCCCGCTAGAGAACAAATTTGCTGTGGGAAGTGGAGCA CATGACTCATTTGTTTGTTACTTTGAGTCTGAAAATGACTGGTGGGTAAGCAAGCATATTAAAAAGCC GATTCGCTCCACAGTCCTCAGCTTGGATTGGCATCCCAACAATGTTTTGCTGGCAGCAGGATCATGTG ACTTCAAATGCAGAGTGTTTTCTGCTTACATTAAAGAAGTGGATGAAAAGCCAGCCAGCATGCCCTGG GGCACCAAGATGCCTTTTGGGCAGCTGATGTCAGAGTTTGGTGGCAGTGGCACCGGTGGCTGGGTCCA CGGGGTAAGCTTCTCTGCCAGTGGGAGCCGCCTGGCCTGGGTCAGCCACGACAGCACCGTTTCAGCCA CGACAGCACCGTTATCTGTTGCTGATGCCTCAAAAAGTGTGCAGGTCTCAATTCTGAAGACAGAGTTC CTACCACTCCTGAGTGTGTCATTTGTCTCAGAGAACAGTGTCGTGGCTGCTGGCCATGACTGCTACCC AATGCTCTTTAATTATGATGACCGCGGCTGCCTGACCTTCGTCTCCAAGTTAGACATTCCAAAACAGA GCATCCAACGCAACATGTCTGCCATAGAACACTTCCGCAACATGGACACGAGGGCCACGATTGAGGAC CACAACATGGCCTTGGAGAGGCTGCACCAGAATAGCATCACTCACGTCTCTATTTATGAGGTAGACAA GCAAGGTTGTCGCAAATTTTGCACTACTGGCATTGATGGAGCCATGACAATTTGGGATTTCAAGACCC TCGAGTCTTCCATCCAGGTCCTCCACATAATGTGAAGCTGAGTGAGCCTTCGCCATCTAGCA		
	ORF Start: ATG at 55		ORF Stop: TGA at 1189
	SEQ ID NO: 38	378 aa	MW at 42383.8kD
NOV13a, CG147651-01 Protein Sequence	MTSLHQFLLEPITCRAWNRDRTQIALSPDNQEAHIYKKHGSQRVKARELSEHNGHITGIHWAPKSDRI VTCGVDCNAYVWSQKDGVWKPTLVILRINRAATFVKWSPLENKFAVGSGAHDSFVCYFESENDWWVSK HIKKPIRSTVLSLDWHPNNVLLAAGSCDFKCRVFSAYIKEVDEKPASMPWGTKMPFGQLMSEFGGSGT GGWVHGVSFSASGSRLAWVSHDSTVSATTAPLSVADASKSVQVSILKTEFLPLLSVSFVSENSVVAAG HDCYPMLFNYDDRGCLTFVSKLDIPKQSIQRNMSAIEHFRNMDTRATIEDHNMALERLHQNSITHVSI YEVDKQGCRKFCTTGIDGAMTIWDFKTLESSIQVLHIM		

Further analysis of the NOV13a protein yielded the following properties shown in Table 13B.

Table 13B. Protein Sequence Properties NOV13a	
PSort analysis:	0.7480 probability located in microbody (peroxisome); 0.4321 probability located in mitochondrial matrix space; 0.1127 probability located in mitochondrial inner membrane; 0.1127 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV13a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13C.

Table 13C. GENESEQ Results for NOV13a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV13a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**

ABG27943	Novel human diagnostic protein #27934 - Homo sapiens, 426 aa. [WO200175067-A2, 11-OCT-2001]	3..378 38..426	329/395 (83%) 335/395 (84%)	0.0
ABG27201	Novel human diagnostic protein #27192 - Homo sapiens, 314 aa. [WO200175067-A2, 11-OCT-2001]	59..378 1..314	290/321 (90%) 292/321 (90%)	e-167
ABG21182	Novel human diagnostic protein #21173 - Homo sapiens, 353 aa. [WO200175067-A2, 11-OCT-2001]	48..370 6..353	276/355 (77%) 283/355 (78%)	e-155
ABG34131	Human ARP2/3 complex 41Kd subunit, P41-ARC - Homo sapiens, 372 aa. [WO200222851-A2, 21-MAR-2002]	5..377 4..371	235/377 (62%) 293/377 (77%)	e-140
AAW67857	Human secreted protein encoded by gene 51 clone HAPNO80 - Homo sapiens, 372 aa. [WO9842738-A1, 01-OCT-1998]	5..377 4..371	235/377 (62%) 293/377 (77%)	e-140

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13D.

Table 13D. Public BLASTP Results for NOV13a

Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99PD4	Suppressor of profilin/p41 of actin-related complex 2/3 - Rattus norvegicus (Rat), 370 aa.	3..378 2..370	335/376 (89%) 342/376 (90%)	0.0
Q9R0Q6	SID329P (Actin related protein 2/3 complex, subunit 1A) (41 kDa) - Mus musculus (Mouse), 370 aa.	3..378 2..370	334/376 (88%) 342/376 (90%)	0.0
Q92747	Actin-related protein 2/3 complex subunit 1A (SOP2-like protein) - Homo sapiens (Human), 370 aa.	3..378 2..370	336/376 (89%) 341/376 (90%)	0.0
Q9BU00	Hypothetical 44.0 kDa protein - Homo sapiens (Human), 401 aa (fragment).	5..377 33..400	235/377 (62%) 293/377 (77%)	e-140
O15143	ARP2/3 complex 41 kDa subunit (P41-ARC) (Actin-related protein 2/3 complex subunit 1B) - Homo sapiens (Human), 372 aa.	5..377 4..371	235/377 (62%) 293/377 (77%)	e-140

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13E.

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	**NOV13a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
WD40	45..81	8/37 (22%) 29/37 (78%)	0.001
WD40	135..171	12/37 (32%) 28/37 (76%)	0.13

Example 14.

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 39	1707 bp	
NOV14a, CG149303-01 DNA Sequence	ACGGATGATGAACAAGCTTTTCATCGGGAACCTGAGCCCCGCCGTCACCGCCGAAGACCTCCGGCAGC TCTTTGGGGACAGGAAGCTGCCCCTGGCGGGACAGGTCCTGCTCAAGTCCCGCTACGCCTTCGTGGAC TACCCCGACCAGAACTGGGCCATCCGCACCATCGAGACCCTCTCGGGTCAAGTGGAATTGCATGGGAA AATCATGGAAGTTGATTATTCAGTCTCTATAAAGCTAAGGAGCAGGAACATTCCGATTCGAAATATCC CTCCTCACCTGCAGTGGGAGGTGTTGGATGGACTTTTGGCTCAATATGGGACAGTGGAGAATGTGGAA CAAGTCAACACAGACACAGAGACCGCTGTTGTCAACGTCACATATGCAACAAAAGAAGAAGTAAAAAT AGCCATGAAGAAGCTAAGCGGGCATCAGTTTGAGAACCACTACTTCAAGATTTCCTACATCCCGGATG ACGAGGTGAGCTGCCCTTCGCCCCCTCAGCGAGCCCAGCGTGGGGACCACTCTTCCTGGGAGCAAGGC CAAGCCCCTGGGGGCTCTTCTCAGGCCAGACAGATTGATTTCCCACTGCGTG1CCTGTTCCCCACCCA GTTTGTTGGTGCCATCATCGGAAAGGAGGGCTTGACCATAAAGAACATCACTAAGCAGAGCCGGTCCC GGGTAGACATCTATAGACAAGAGAACTCCAGAGCTGCAGAGAAGCCTGTCACCATGCATGCCACCCCA GAGGGGACTTCTGAAGCATGCCGCATGATTCTTGAAATAATGCAGAAAGAGGCAGATGAGGCCAAACT AGCCGAAGAGATTCCTCTGAAAATCTTGGCCCACAATGGCTTGGTTGGAAGACTGATTGGAAAAGAAG GCAGAAATTTGAAGAAAAATGAACATGAAACAGGGACCAAGATAACAATCTCATCTTCGCAGGATTTG AGCATATACAACCCGGAAAGAACCATCACTGTGAAGGGCACAGTCGAGGTCTGTGCCAGTGCTGAGAT AGAGATTATGAAGAAGCTGCGTGAGGCCTTTGAAAATGATACGCTGACTGTTAATACCCACTTCGGAT ACTTCTCCAGCCTGTACCCCCATCGCCAGTTTGGCCCGTTCCCGCATCATCACTCTTATCCAGAGCAG GAGATTGTCAATCTCTTCATCCCAACCCAGGGTGTGGGCGCCATCATCGGGAAGAAAGGGGCACACAT CAAACAGCTGGCGAGATTCGTGGGAGCCTCCATCAAGATCGCCCCTGCGAGATCGCCCCTGCGTCAGC GGAAGGTCATCATCACCTGGCCACCGGAATCCCAGTTCAAGGCCCAGGGACGGATCTTTGGGAAACTG AAAGAAGAAAACTTTTTTAACCCCAAAGAAGACGTGAAGCTGGAAACCCATATCAGAGTGCCCTCTTC CACCGCTGGCCGGGTGATTGGCAAAGGGGGCAAGACCGTGAATGAACTGCAGAATTTAATCAGTGCAG AAGTCATCGTGCCTCGTGACCAAACGCCAGATGAAAATGAGGAAATGATCGTCAGAATTATCGGGCAC TTCTTTGCTAGCCAGACTGCACAGCGCAAGATCAGGGAAATTGTACAACAGGTGAAGCAGCAGGAGCA GAAATACCCTCAGGGAGTCGCCTCACAGCGCAGCAAGTGAGGATCCCACAGGCACAAGCAAAACAACG GAAGAAT		
	ORF Start: ATG at 5		ORF Stop: TGA at 1670
	SEQ ID NO: 40	555 aa	MW at 62450.8kD
NOV14a, CG149303-01 Protein Sequence	MMNKLFIGNLSPAVTAEDLRQLFGDRKLPLAGQVLLKSRYAFVDYPDQNWAIRTIETLSGQVELHGKI MEVDYSVSIKLRSRNIPIRNIPPHLQWEVLDGLLAQYGTVENVEQVNTDTETAVVNVTYATKEEVKIA MKKLSGHQFENHYFKISYIPDDEVSCPSPPQRAQRGDHSSWEQGQAPGGSSQARQIDFPLRVLFPTQF VGAIIGKEGLTIKNITKQSRSRVDIYRQENSRAAEKPVTMHATPEGTSEACRMILEIMQKEADEAKLA EEIPLKILAHNGLVGRLIGKEGRNLKKNEHETGTKITISSSQDLSIYNPERTITVKGTVEVCASAEIE IMKKLREAFENDTLTVNTHFGYFSSLYPHRQFGPFPHHHSYPEQEIVNLFIPTQGVGAIIGKKGAHIK QLARFVGASIKIAPARSPLRQRKVIITWPPESQFKAQGRIFGKLKEENFFNPKEDVKLETHIRVPSST AGRVIGKGGKTVNELQNLISAEVIVPRDQTPDENEEMIVRIIGHFFASQTAQRKIREIVQQVKQQEQK YPQGVASQRSK		

Further analysis of the NOV14a protein yielded the following properties shown in Table 14B.

Table 14B. Protein Sequence Properties NOV14a	
PSort analysis:	0.5050 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV14a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14C.

Table 14C. GENESEQ Results for NOV14a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV14a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ABG06795	Novel human diagnostic protein #6786 - Homo sapiens, 594 aa. [WO200175067-A2, 11-OCT-2001]	1..555 39..594	507/556 (91%) 526/556 (94%)	0.0
AAU16163	Human novel secreted protein, Seq ID 1116 - Homo sapiens, 620 aa. [WO200155322-A2, 02-AUG-2001]	1..555 22..620	507/599 (84%) 526/599 (87%)	0.0
ABG06794	Novel human diagnostic protein #6785 - Homo sapiens, 614 aa. [WO200175067-A2, 11-OCT-2001]	1..555 16..614	507/599 (84%) 526/599 (87%)	0.0
AAY30649	A murine c-myc coding region determinant binding protein - Mus musculus, 577 aa. [WO9946594-A2, 16-SEP-1999]	2..555 1..577	353/579 (60%) 442/579 (75%)	0.0
ABB75054	Human lung tumour L523S recombinant protein sequence SEQ ID NO:449 - Homo sapiens, 579 aa. [WO200200174-A2, 03-JAN-2002]	2..544 1..566	343/567 (60%) 424/567 (74%)	0.0

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14D.

Table 14D. Public BLASTP Results for NOV14a				
Protein Accession Number	**Protein/Organism/Length**	**NOV14a Residues/ Match Residues**	**Identities/ Similarities for the Matched Portion**	**Expect Value**
Q9Y6M1	Hepatocellular carcinoma autoantigen - Homo sapiens (Human), 556 aa.	1..555 1..556	507/556 (91%) 526/556 (94%)	0.0
AAD09827	IGF-II MRNA-BINDING PROTEIN 2 - Homo sapiens (Human), 598 aa.	2..555 1..598	506/598 (84%) 525/598 (87%)	0.0
O42254	Zipcode-binding protein - Gallus gallus (Chicken), 576 aa.	2..555 1..576	360/583 (61%) 440/583 (74%)	0.0
AAD09826	IGF-II MRNA-BINDING PROTEIN 1 - Homo sapiens (Human), 577 aa.	2..555 1..577	355/579 (61%) 442/579 (76%)	0.0
O73932	VG1 RNA binding protein variant D - Xenopus laevis (African clawed frog), 594 aa.	2..551 1..586	355/586 (60%) 440/586 (74%)	0.0

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14E.

Table 14E. Domain Analysis of NOV14a			
Pfam Domain	**NOV14a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
rrm	5..71	24/78 (31%) 46/78 (59%)	9.3e-05
rrm	84..152	14/77 (18%) 48/77 (62%)	7.3e-05
KH-domain	197..244	17/49 (35%) 38/49 (78%)	3.7e-06
KH-domain	278..329	19/52 (37%) 41/52 (79%)	1.9e-07
KH-domain	388..436	18/49 (37%) 33/49 (67%)	4.9e-05
KH-domain	469..519	18/51 (35%) 42/51 (82%)	3.5e-08

Example 15.

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis			
	SEQ ID NO: 41	400 bp	
NOV15a, CG149312-01 DNA Sequence	CGACGGCGCCATGAGTCTGACTTCCAGTTCCAGCGGACGAGTTGAATGGATGGCAGCAGTTACCGTTG CTGCTGGGACAGCTGCAATTGGTTATCTAGCTTACAAAAGATTTTATGTTAAAGATCATCGAAATAAA GCTGTGATAAACCTTCACATCCAGAAAGACAACCCCAAGACAGTACATGCTTTTGACATGGAGGATTT GGGAGATAATGCTGTGTACTGCCGTTTCTGGAGGTCCAAAAATTCCCATTCTGTGATGGGTCTCACAC AAAACACAACGAAGAGACTGGAGTCAACGTGGGACAAATTCCCATTCTGTGATGGGTCTCACACAAAA CACAACGAAGAGACTGGAGTCAACGTGGGACCTCTTATCATCAAGAAAAAAGAAACTTAA		
	ORF Start: ATG at 11		ORF Stop: TAA at 398
	SEQ ID NO: 42	129 aa	MW at 14534.4kD
NOV15a, CG149312-01 Protein Sequence	MSLTSSSSGRVEWMAAVTVAAGTAAIGYLAYKRFYVKDHRNKAVINLHIQKDNPKTVHAFDMEDLGDN AVYCRFWRSKNSHSVMGLTQNTTKRLESTWDKFPFCDGSHTKHNEETGVNVGPLIIKKKET		

Further analysis of the NOV15a protein yielded the following properties shown in Table 15B.

Table 15B. Protein Sequence Properties NOV15a	
PSort analysis:	0.8598 probability located in mitochondrial intermembrane space; 0.7605 probability located in mitochondrial matrix space; 0.4691 probability located in microbody (peroxisome); 0.4392 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV15a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15C.

Table 15C. GENESEQ Results for NOV15a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV15a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAY36115	Extended human secreted protein sequence, SEQ ID NO. 500 - Homo sapiens, 108 aa. [WO9931236-A2, 24-JUN-1999]	1..129 1..108	99/129 (76%) 103/129 (79%)	1e-50

AAW64556	Human osterocarcinoma cell line U-2 OS clone HP10305 protein - Homo sapiens, 108 aa. [WO9821328-A2, 22-MAY-1998]	1..129 1..108	99/129 (76%) 103/129 (79%)	1e-50
AAY12490	Human 5' EST secreted protein SEQ ID NO:521 - Homo sapiens, 96 aa. [WO9906548-A2, 11-FEB-1999]	1..117 1..96	83/117 (70%) 88/117 (74%)	7e-41
ABG17140	Novel human diagnostic protein #17131 - Homo sapiens, 395 aa. [WO200175067-A2, 11-OCT-2001]	10..129 297..395	80/120 (66%) 87/120 (71%)	3e-40
ABG17141	Novel human diagnostic protein #17132 - Homo sapiens, 105 aa. [WO200175067-A2, 11-OCT-2001]	22..129 19..105	78/108 (72%) 81/108 (74%)	9e-39

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15D.

Table 15D. Public BLASTP Results for NOV15a

Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NZ45	Uncharacterized hematopoietic stem/progenitor cells protein MDS029 - Homo sapiens (Human), 108 aa.	1..129 1..108	99/129 (76%) 103/129 (79%)	3e-50
Q8WUQ5	Hypothetical 12.1 kDa protein - Homo sapiens (Human), 108 aa.	1..129 1..108	86/129 (66%) 97/129 (74%)	2e-43
Q9D0Y0	1500009M05Rik protein - Mus musculus (Mouse), 135 aa.	12..128 37..134	53/119 (44%) 67/119 (55%)	3e-19
Q9CQB5	1500009M05Rik protein - Mus musculus (Mouse), 135 aa.	12..128 37..134	53/119 (44%) 67/119 (55%)	5e-19
AAH32300	Similar to RIKEN cDNA 1500009M05 gene - Homo sapiens (Human), 135 aa.	12..128 37..134	54/119 (45%) 66/119 (55%)	7e-19

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15E.

Table 15E. Domain Analysis of NOV15a

Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found To Publicly Searchable Domains			

Example 16.

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

5

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 43	786 bp	
NOV16a, CG150951-01 DNA Sequence	CGACAGAAGGGTACGGCTGCCAGAAGACGACAGAAGGGTACGGCTGCGAGAAGACGACAGAAGGGTAC GGCTGCGAGAAGACGACAGAAGGGGGCTCTTCCTCGTTTGCCCCTCGTGTTCATGGGAGCTCGTTTTC TTTTCCTCTAGGCAGAGAAGAGGCGATGGCGGCGATGGCATCTCTCGGCGCCCTGGCGCTGCTCCTGC TGTCCAGCCTCTCCCGCTGCTCAGCCGAGGCCTGCCTGGAGCCCCAGATCACCCCTTCCTACTACACC ACTTCTGACGCTGTCATTTCCACTGAGACCGTCTTCATTGTGGAGATCTCCCTGACATGCAAGAACAG GGTCCAGAACATGGCTCTCTATGCTGACGTCGGTGGAAAACAATTCCCTGTCACTCGAGGCCAGGATG TGGGGCGTTATCAGGTGTCCTGGAGCCTGGACCACAAGAGCGCCCACGCAGGCACCTATGAGGTTAGA TTCTTCGACGAGGAGTCCTACAGCCTCCTCAGGAAGGCTCAGAGGAATAACGAGGACATTTCCATCAT CCCGCCTCTGTTTACAGTCAGCGTGGACCATCGGGGCACTTGGAACGGGCCCTGGGTGTCCACTGAGG TGCTGGCTGCGGCGATCGGCCTTGTGATCTACTACTTGGCCTTCAGTGCGAAGAGCCACATCCAGGCC TGAGGGCGGCACCCCAGCCCTGCCCTTGCTTCCTTCAATAAACATCACAGGACCTGGGACTGCACAGG AAAAAAAAAAAAAAACTCGNGGGGGGCCCGGTACCCAA		
	ORF Start: ATG at 162		ORF Stop: TGA at 681
	SEQ ID NO: 44	173 aa	MW at 18998.4kD
NOV16a, CG150951-01 Protein Sequence	MAAMASLGALALLLLSSLSRCSAEACLEPQITPSYYTTSDAVISTETVFIVEISLTCKNRVQNMALYA DVGGKQFPVTRGQDVGRYQVSWSLDHKSAHAGTYEVRFFDEESYSLLRKAQRNNEDISIIPPLFTVSV DHRGTWNGPWVSTEVLAAAIGLVIYYLAFSAKSHIQA		
	SEQ ID NO: 45	623 bp	
NOV16b, CG150951-02 DNA Sequence	CGACAGAAGGGTACGGCTGCCAGAAGACGACAGAAGGGTACGGCTGCGAGAAGACGACAGAAGGGTAC GGCTGCGAGAAGACGACAGAAGGGGGCTCTTCCTCGTTTGCCCCTCGTGTTCATGGGAGCTCGTTTTC TTTTCCTCTAGGCAGAGAAGAGGCGATGGCGGCGATGGCATCTCTCGGCGCCCTGGCGCTGCTCCTGC TGTCCAGCCTCTCCCGCTGCTCAGCCGAGGCCTGCCTGGAGCCCCAGATCACCCCTTCCTACTACACC ACTTCTGACGCTGTCATTTCCACTGAGACCGTCTTCATTGTGGAGATCTCCCTGACATGCAAGAACAG GGTCCAGGTGTCCTGGAGCCTGGACCACAAGAGCGCCCACGCAGGCACCTATGAGGTTAGATTCTTCG ACGAGGAGTCCTACAGCCTCCTCAGGAAGGCTCAGAGGAATAACGAGGACATTTCCATCATCCCGCCT CTGTTTACAGTCAGCGTGGACCATCGGGGCACTTGGAACGGGCCCTGGGTGTCCACTGAGGTGCTGGC TGCGGCGATCGGCCTTGTGATCTACTACTTGGCCTTCAGTGCGAAGAGCCACATCCAGGCCTGAGGAA GGGCGAATTCC		
	ORF Start: ATG at 162		ORF Stop: TGA at 606
	SEQ ID NO: 46	148 aa	MW at 16245.3kD
NOV16b, CG150951-02 Protein Sequence	MAAMASLGALALLLLSSLSRCSAEACLEPQITPSYYTTSDAVISTETVFIVEISLTCKNRVQVSWSLD HKSAHAGTYEVRFFDEESYSLLRKAQRNNEDISIIPPLFTVSVDHRGTWNGPWVSTEVLAAAIGLVIY YLAFSAKSHIQA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 16B.

Table 16B. Comparison of NOV16a against NOV16b.		
Protein Sequence	**NOV16a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**

NOV16b	26..173 26..148	123/148 (83%) 123/148 (83%)

Further analysis of the NOV16a protein yielded the following properties shown in Table 16C.

Table 16C. Protein Sequence Properties NOV16a

PSort analysis:	0.9190 probability located in plasma membrane; 0.2000 probability located in lysosome (membrane); 0.1339 probability located in microbody (peroxisome); 0.1000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	Cleavage site between residues 24 and 25

A search of the NOV16a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16D.

Table 16D. GENESEQ Results for NOV16a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG75579	Human colon cancer antigen protein SEQ ID NO:6343 - Homo sapiens, 226 aa. [WO200122920-A2, 05-APR-2001]	1..173 54..226	173/173 (100%) 173/173 (100%)	1e-95
AAB43566	Human cancer associated protein sequence SEQ ID NO:1011 - Homo sapiens, 187 aa. [WO200055350-A1, 21-SEP-2000]	1..173 15..187	173/173 (100%) 173/173 (100%)	1e-95
ABP42089	Human ovarian antigen HAZAG23, SEQ ID NO:3221 - Homo sapiens, 155 aa. [WO200200677-A1, 03-JAN-2002]	19..173 1..155	149/155 (96%) 150/155 (96%)	1e-82
AAB87645	Bovine mammary tissue derived protein #36 - Bos taurus, 152 aa. [WO200114553-A1, 01-MAR-2001]	4..155 1..152	145/152 (95%) 149/152 (97%)	1e-80
AAB87646	Bovine mammary tissue derived protein #37 - Bos taurus, 105 aa. [WO200114553-A1, 01-MAR-2001]	77..173 9..105	95/97 (97%) 97/97 (99%)	2e-51

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16E.

Table 16E. Public BLASTP Results for NOV16a

Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P51571	Translocon-associated protein, delta subunit precursor (TRAP-delta) (Signal sequence receptor delta subunit) (SSR-delta) - Homo sapiens (Human), 173 aa.	1..173 1..173	173/173 (100%) 173/173 (100%)	4e-95
Q07984	Translocon-associated protein, delta subunit precursor (TRAP-delta) (Signal sequence receptor delta subunit) (SSR-delta) - Rattus norvegicus (Rat), 173 aa.	1..173 1..173	166/173 (95%) 169/173 (96%)	9e-92
Q9D8L3	Signal sequence receptor, delta - Mus musculus (Mouse), 173 aa.	1..173 1..173	165/173 (95%) 168/173 (96%)	3e-91
Q9DC94	Signal sequence receptor, delta - Mus musculus (Mouse), 173 aa.	1..173 1..173	164/173 (94%) 168/173 (96%)	8e-91
Q62186	Translocon-associated protein, delta subunit precursor (TRAP-delta) (Signal sequence receptor delta subunit) (SSR-delta) - Mus musculus (Mouse), 172 aa.	1..173 1..172	164/173 (94%) 167/173 (95%)	2e-89

PFam analysis predicts that the NOV16a protein contains the domains shown in the Table 16F.

Table 16F. Domain Analysis of NOV16a

Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
No Significant Matches Found To Publicly Searchable Domains			

Example 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis

	SEQ ID NO: 47	1831 bp
NOV17a, CG173328-01 DNA Sequence	GCTTGACTCAGACCAAGTCCCATCACTGGTCGCACTGACGGGTGAAGTGGCAGGAGGAGGAGGAGGGG CTCCGGCTGGTCTGTGGTGAGATGGCCTACCAGGTGGTGGAGAAGGGCGCGGCCCTGGGCACGCTGGA GTCGGAGCTGCAGCAGAGGCAAAGCAGGCTGGCAGCCCTGGAGGCCCGCGTGGCGCAGCTGCGAGAGG	

	CGCGGGCGCAGCAGGCCCAGCAGGTGGAGGAGTGGCGGGCGCAGAATGCGGTGCAGCGGGCAGCCTAC GAGGCGCTGCGCGCGCACGTCGGGCTCCGGGAGGCGGCACTGCGCAGGCTCCAGGAAGAGGCGCGCGA CCTGCTGGAGAGGCTCGTGCAGCGCAAGGCGCGCGCCGCGGCCGAGCGCAACCTGCGCAACGAGCGCC GGGAGCGGGCCAAGCAGGCGCGGGTGTCCCAGGAGCTGAAGAAGGCTGCCAAGCGGACCGTGAGCATC AGCGAGGGCCCGGACACCCTAGGCGATGGGATGAGGGAGAGAAGGGAGACTCTGGCTCTGGCCCCTGA GCCAGAGCCCCTGGAGAAGGAAGCTTGTGAGAAGTGGAAGAGGCCCTTCAGTTTTAAGAAGAGGAGAG GTCACTCAATTGGGGGAGCCCCTGAGCAGCGATACCAGATCATCCCTGTGTGTGTGGCTGCCCGACTT CCTACCCGGGCTCAGGATGTGCTGGATGCCCACCTCTCTGAGGTCAATGCTGTTCGTTTTGGCCCCAA CAGCAGCCTCCTGGCCACTGGAGGGGCTGACCGCCTGATCCACCTCTGGAATGTTGTGGGAAGTCGCC TGGAGGCCAACCAGACCCTGGAGGGAGCTGGTGGCAGCATCACCAGTGTGGACTTTGACCCCTCGGGC TACCAGGTTTTAGCAGCAACTTACAACCAGGCTGCCCAGCTCTGGAAGGTGGGGGAGGCACAGTCCAA GGAGACACTGTCTGGACACAAGGATAAGGTGACAGCTGCCAAATTCAAGCTAACGAGGCACCAGGCAG TGACTGGGAGCCGCGACCGGACAGTGAAGGAGTGGGACCTCGGCCGTGCCTATTGCTCCAGGACCATC AATGTCCTTTCCTACTGTAATGACGTGGTGTGTGGGGACCATATCATCATTAGTGGCCACAATGACCA GAAGATCCGGTTCTGGGACAGCAGGGGGCCCCACTGCACCCAGGTCATCCCTGTGCAGGGCCGGGTCA CCTCCCTGAGCCTCAGCCACGACCAACTGCACCTGCTCAGCTGTTCCCGAGACAACACACTCAAGGTC ATCGACCTGCGTGTCAGCAACATCCGCCAGGTGTTCAGGGCCGATGGCTTCAAGTGTGGTTCTGACTG GACCAAAGCTGTGTTCAGCCCGGACAGAAGCTATGCACTGGCAGGCTCCTGTGATGGGGCCCTTTACA TCTGGGATGTGGACACCGGGAAACTGGAGAGCAGACTACAGGGACCCCATTGCGCTGCCGTCAACGCC GTGGCCTGGTGCTACTCCGGGAGCCACATGGTGAGCGTGGACCAGGGCAGGAAGGTTGTGCTCTGGCA GTAGGGCCACGACCTGCCTGCCTGGGCTGGAGCTCTTGCCCGAAGCCTGAAGCTTCCTTCGGCGCCAT GCAGGGGTTGGGGTTGGGACTGGAGCTGGCCTTGGGATTTAATGGGGAAGAAGGCCTGGCAGGACCTG GCCTGTTTGTTTAAAAATGAAGTATGGGTTGGGGGATTACGCTAGTTTTTCTTTGTATTTTTATCTCT ATCTATCTCCTCACTTTTTCTCCCAAAGTAGAAAAAAATGATATCTGAAAAAAAAAAAAAAAAA
	ORF Start: ATG at 90 | ORF Stop: TAG at 1566
	SEQ ID NO: 48 | 492 aa | MW at 54786.4kD
NOV17a, CG173328-01 Protein Sequence	MAYQVVEKGAALGTLESELQQRQSRLAALEARVAQLREARAQQAQQVEEWRAQNAVQRAAYEALRAHV GLREAALRRLQEEARDLLERLVQRKARAAAERNLRNERRERAKQARVSQELKKAAKRTVSISEGPDTL GDGMRERRETLALAPEPEPLEKEACEKWKRPFSFKKRRGHSIGGAPEQRYQIIPVCVAARLPTRAQDV LDAHLSEVNAVRFGPNSSLLATGGADRLIHLWNVVGSRLEANQTLEGAGGSITSVDFDPSGYQVLAAT YNQAAQLWKVGEAQSKETLSGHKDKVTAAKFKLTRHQAVTGSRDRTVKEWDLGRAYCSRTINVLSYCN DVVCGDHIIISGHNDQKIRFWDSRGPHCTQVIPVQGRVTSLSLSHDQLHLLSCSRDNTLKVIDLRVSN IRQVFRADGFKCGSDWTKAVFSPDRSYALAGSCDGALYIWDVDTGKLESRLQGPHCAAVNAVAWCYSG SHMVSVDQGRKVVLWQ
	SEQ ID NO: 49 | 1894 bp |
NOV17b, CG173328-02 DNA Sequence	GCTTGACTCAGACCAAGTCCCATCACTGGTCGCACTGACGGGTGAAGTGGCAGGAGGAGGAGGAGGGG CTCCGGCTGGTCTGTGGTGAGATGGCCTACCAGGTGGTGGAGAAGGGCGCGGCCCTGGGCACGCTGGA GTCGGAGCTGCAGCAGAGGCAAAGCAGGCTGGCAGCCCTGGAGGCCCGCGTGGCGCAGCTGCGAGAGG CGCGGGCGCAGCAGGCCCAGCAGGTGGAGGAGTGGCGGGCGCAGAATGCGGTGCAGCGGGCAGCCTAC GAGGCGCTGCGCGCGCACGTCGGGCTCCGGGAGGCGGCACTGCGCAGGCTCCAGGAAGAGGCGCGCGA CCTGCTGGAGAGGCTCGTGCAGCGCAAGGCGCGCGCCGCGGCCGAGCGCAACCTGCGCAACGAGCGCC GGGAGCGGGCCAAGCAGGCGCGGGTGTCCCAGGAGCTGAAGAAGGCTGCCAAGCGGACCGTGAGCATC AGCGAGGGCCCGGACACCCTAGGCGATGGGATGAGGGAGAGAAGGGAGACTCTGGCTCTGGCCCCTGA GCCAGAGCCCCTGGAGAAGGAAGCTTGTGAGAAGTGGAAGAGGCCCTTCAGGTCTGCCTCAGCCACCT CCCTGACGCTGTCCCACTGTGTGGATGTGGTGAAGGGGCTTCTGGATTTTAAGAAGAGGAGAGGTCAC TCAATTGGGGGAGCCCCTGAGCAGCGATACCAGATCATCCCTGTGTGTGTGGCTGCCCGACTTCCTAC CCGGGCTCAGGATGTGCTGGATGCCCACCTCTCTGAGGTCAATGCTGTTCGTTTTGGCCCCAACAGCA GCCTCCTGGCCACTGGAGGGGCTGACCGCCTGATCCACCTCTGGAATGTTGTGGGAAGTCGCCTGGAG GCCAACCAGACCCTGGAGGGAGCTGGTGGCAGCATCACCAGTGTGGACTTTGACCCCTCGGGCTACCA GGTTTTAGCAGCAACTTACAACCAGGCTGCCCAGCTCTGGAAGGTGGGGGAGGCACAGTCCAAGGAGA CACTGTCTGGACACAAGGATAAGGTGACAGCTGCCAAATTCAAGCTAACGAGGCACCAGGCAGTGACT GGGAGCCGCGACCGGACAGTGAAGGAGTGGGACCTCGGCCGTGCCTATTGCTCCAGGACCATCAATGT CCTTTCCTACTGTAATGACGTGGTGTGTGGGGACCATATCATCATTAGTGGCCACAATGACCAGAAGA TCCGGTTCTGGGACAGCAGGGGGCCCCACTGCACCCAGGTCATCCCTGTGCAGGGCCGGGTCACCTCC CTGAGCCTCAGCCACGACCAACTGCACCTGCTCAGCTGTTCCCGAGACAACACACTCAAGGTCATCGA CCTGCGTGTCAGCAACATCCGCCAGGTGTTCAGGGCCGATGGCTTCAAGTGTGGTTCTGACTGGACCA AAGCTGTGTTCAGCCCGGACAGAAGCTATGCACTGGCAGGCTCCTGTGATGGGGCCCTTTACATCTGG GATGTGGACACCGGGAAACTGGAGAGCAGACTACAGGGACCCCATTGCGCTGCCGTCAACGCCGTGGC CTGGTGCTACTCCGGGAGCCACATGGTGAGCGTGGACCAGGGCAGGAAGGTTGTGCTCTGGCAGTAGG GCCACGACCTGCCTGCCTGGGCTGGAGCTCTTGCCCGAAGCCTGAAGCTTCCTTCGGCGCCATGCAGG GGTTGGGGTTGGGACTGGAGCTGGCCTTGGGATTTAATGGGGAAGAAGGCCTGGCAGGACCTGGCCTG TTTGTTTAAAAATGAAGTATGGGTTGGGGGATTACGCTAGTTTTTCTTTGTATTTTTATCTCTATCTA TCTCCTCACTTTTTCTCCCAAAGTAGAAAAAAATGATATCTGAAAAAAAAAAAAAAAAA
	ORF Start: ATG at 90 | ORF Stop: TAG at 1629

	SEQ ID NO: 50	513 aa	MW at 56953.9kD
NOV17b, CG173328-02 Protein Sequence	MAYQVVEKGAALGTLESELQQRQSRLAALEARVAQLREARAQQAQQVEEWRAQNAVQRAAYEALRAHV GLREAALRRLQEEARDLLERLVQRKARAAAERNLRNERRERAKQARVSQELKKAAKRTVSISEGPDTL GDGMRERRETLALAPEPEPLEKEACEKWKRPFRSASATSLTLSHCVDVVKGLLDFKKRRGHSIGGAPE QRYQIIPVCVAARLPTRAQDVLDAHLSEVNAVRFGPNSSLLATGGADRLIHLWNVVGSRLEANQTLEG AGGSITSVDFDPSGYQVLAATYNQAAQLWKVGEAQSKETLSGHKDKVTAAKFKLTRHQAVTGSRDRTV KEWDLGRAYCSRTINVLSYCNDVVCGDHIIISGHNDQKIRFWDSRGPHCTQVIPVQGRVTSLSLSHDQ LHLLSCSRDNTLKVIDLRVSNIRQVFRADGFKCGSDWTKAVFSPDRSYALAGSCDGALYIWDVDTGKL ESRLQGPHCAAVNAVAWCYSGSHMVSVDQGRKVVLWQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 17B.

Table 17B. Comparison of NOV17a against NOV17b.

Protein Sequence	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV17b	1..492 1..513	409/513 (79%) 409/513 (79%)

Further analysis of the NOV17a protein yielded the following properties shown in Table 17C.

Table 17C. Protein Sequence Properties NOV17a

PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1500 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV17a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17D.

Table 17D. GENESEQ Results for NOV17a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE23378	Human intracellular signaling (INTSIG-5) protein - Homo sapiens, 513 aa. [WO200231152-A2, 18-APR-2002]	1..492 1..513	491/513 (95%) 491/513 (95%)	0.0

AAE23374	Human intracellular signaling (INTSIG-1) protein - Homo sapiens, 435 aa. [WO200231152-A2, 18-APR-2002]	1..414 1..435	413/435 (94%) 413/435 (94%)	0.0
AAM41955	Human polypeptide SEQ ID NO 6886 - Homo sapiens, 446 aa. [WO200153312-A1, 26-JUL-2001]	109..492 42..446	383/405 (94%) 383/405 (94%)	0.0
AAM40169	Human polypeptide SEQ ID NO 3314 - Homo sapiens, 359 aa. [WO200153312-A1, 26-JUL-2001]	140..482 1..356	328/364 (90%) 329/364 (90%)	0.0
ABG07689	Novel human diagnostic protein #7680 - Homo sapiens, 429 aa. [WO200175067-A2, 11-OCT-2001]	112..487 1..407	327/407 (80%) 332/407 (81%)	e-176

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17E.

Table 17E. Public BLASTP Results for NOV17a

Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAC04021	CDNA FLJ35698 fis, clone SPLEN2019839, weakly similar to TIPD PROTEIN - Homo sapiens (Human), 619 aa.	1..492 107..619	490/513 (95%) 490/513 (95%)	0.0
BAC03485	CDNA FLJ33278 fis, clone ASTRO2008508, weakly similar to TIPD PROTEIN - Homo sapiens (Human), 513 aa.	1..492 1..513	489/513 (95%) 489/513 (95%)	0.0
Q96D26	Hypothetical 52.5 kDa protein - Homo sapiens (Human), 472 aa.	16..491 5..469	190/478 (39%) 278/478 (57%)	2e-95
Q9BR11	Hypothetical 52.6 kDa protein - Homo sapiens (Human), 472 aa.	16..491 5..469	190/478 (39%) 278/478 (57%)	2e-95
Q96JV5	CDNA FLJ14948 fis, clone PLACE2000164, weakly similar to TIPD protein - Homo sapiens (Human), 504 aa.	1..491 22..501	191/493 (38%) 283/493 (56%)	3e-95

PFam analysis predicts that the NOV17a protein contains the domains shown in the Table 17F.

Table 17F. Domain Analysis of NOV17a			
Pfam Domain	**NOV17a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
WD40	201..237	14/37 (38%) 30/37 (81%)	0.0003
WD40	287..323	15/37 (41%) 26/37 (70%)	0.0022
WD40	368..403	10/37 (27%) 33/37 (89%)	0.0068

Example 18.

The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis			
	SEQ ID NO: 51	622 bp	
NOV18a, CG56101-01 DNA Sequence	CTCCTTTCCAAAGAACCATGAGTTCCCACATCAGCCAGAATTACTGCACCGAAGTGGAAGCCGCCGTC AGCAGCCTGGTCCACCGGCAGCTGCGGGCTTCCCTTACCTACCTCTCTCTCATCCTCCATTTCTACCG CGACGACGTGACCCTGGAGGGCATGGGCCACTTCCGAGAGCTGGCCCAGGAGAAGCGACAGGGCGCCC AGAGTCTGTGGAAGACGCAAAACCAGCGCGGAGCCCTCTGCGATGCCATCCAGAAGCCGTCCTGGGAT GAAAAGGACAGCAGTTTGGGCGCCCTGCGAGCCGCGTTGGCCCTGGAGACGAACCTGAACCAGGCCCT GCTGGATCTGCACGCCCTGGGCGCAAAGCATGCAGACTCTCACCCCTGCGGCTTCCTGGAGAACCACT TCCGGCCACATCCCTCTGTCAGACCTGGGAAAGCGTCCACCCGAGCTGCTCCCTTCAACCTCAAGATA CATTTTTTTTCTTTCTTTCTTTTTGAAAGAGTCTCCCTGCGTGTAGACCCCTGGACTATTGATTGCAC CACATTCATTCCTTCCCCAGCTCACTACTCCAACAAGGTACCAAATATACCAAATATTTAGAGAATTA GGATGAACTA		
	ORF Start: ATG at 18		ORF Stop: TAG at 603
	SEQ ID NO: 52	195 aa	MW at 21962.6kD
NOV18a, CG56101-01 Protein Sequence	MSSHISQNYCTEVEAAVSSLVHRQLRASLTYLSLILHFYRDDVTLEGMGHFRELAQEKRQGAQSLWKT QNQRGALCDAIQKPSWDEKDSSLGALRAALALETNLNQALLDLHALGAKHADSHPCGFLENHFRPHPS VRPGKASTRAAPFNLKIHFFSFFLFERVSLRVDPWTIDCTTFIPSPAHYSNKVPNIPNI		
	SEQ ID NO: 53	502 bp	
NOV18b, CG56101-03 DNA Sequence	CGGTCCCGCGGGTCTGTCTCTTGCTTCAACAGTGTTTGGACGGAACAGATCCGGGGACTCTCTTCCAG CCTCCGACCGCCCTCCGATTTCCTCTCTTCCAGGACATCAAGAAGCCAGCTGAAGATGCAAAACCAGC GTGGCGGCCGCGCTCTCTTCTCGGCCATCTCCTGCTTCTGGGACCTGCCAGCACCGTTTTTGTGGTTA GCTCCTTCTTGCCAACCAACCATGAGCTCCCAGATTCGTCAGAATTATTCCACCGACGTGGAGGCAGC CGTCAACAGCCTGGTCAATTTGTACCTGCAGGCCTCCTACACCTACCTCTCTCTGGGCTTCTATTTCG ACCGCGATGATGTGGCTCTGGAAGGCGTGAGCCACTTCTTCCGCGAATTGGCCGAGGAGAAGCGCGAG GGCTACGAGCGTCTCCTGAAGATGCAAAACCAGCGTGGGGTAAAACCCCAGACGCCATGAAAGCTGCC ATGGCCCTGGAGAAAAAGCTGAACCC		
	ORF Start: ATG at 124		ORF Stop: TGA at 466
	SEQ ID NO: 54	114 aa	MW at 13105.7kD
NOV18b, CG56101-03 Protein Sequence	MQNQRGGRALFSAISCFWDLPAPFLWLAPSCQPTMSSQIRQNYSTDVEAAVNSLVNLYLQASYTYLSL GFYFDRDDVALEGVSHFFRELAEEKREGYERLLKMQNQRGVKPQTP		

	SEQ ID NO: 55	723 bp	
NOV18c, CG56101-02 DNA Sequence	AGTTGTTGCTTATGATGTGTGAGTGAACATATGCCATGCCTGGCCTTTTTTGTGGTTAGCTCCTTCTT GCCAACCAACCATGAGCTCCCAGATTCGTCAGAATTATTCCACCGACGTGGAGGCAGCCGTCAACAGC CTGGTCAATTTGTACCTGCAGGCCTCCTACACCTACCTCTCTCTGGGCTTCTATTTCGACCGCGATGA TGTGGCTCTGGAAGGCGTGAGCCACTTCTTCCGCGAACTGGCCGAGGAGAAGCGCGAGGGCTACGAGC GTCTCCTGAAGATGCAAAACCAGCGTGGCGGCCGCGCTCTCTTCCAGGACATCAAGAAGCCAGCTGAA GATGAGTGGGGTAAAACCCCAGACGCCATGAAAGCTGCCATGACCCTGGAGAAAAAGCTGAACCAGGC CCTTTTGGATCTTCATGCCCTGGGTTCTGCCCGCACGGACCCCCATCTCTGTGACTTCCTGGAGACTC ACTTCCTAGATGAGGAAGTGAAGCTTATCAAGAAGATGGGTGACCACCTGACCAACCTCCACAGGCTG GGTGGCCCCGGAGGCTGGGCTGGGCGAGTATCTCTTCGAAAGGCTCACTCTCAAGCACGACTAAGAGCC TTCTGAGCCCAGCGACTTCTGAAGGGCCCCTTGCAAAGTAATAGGGCTTCTGCCTAAGCCTCTCCCTC CAGCCAATAGGCAGCTTTCTTAACTATCCTAACAAGCCTTGGA		
	ORF Start: ATG at 80		ORF Stop: TAA at 605
	SEQ ID NO: 56	175 aa	MW at 20049.5kD
NOV18c, CG56101-02 Protein Sequence	MSSQIRQNYSTDVEAAVNSLVNLYLQASYTYLSLGFYFDRDDVALEGVSHFFRELAEEKREGYERLLK MQNQRGGRALFQDIKKPAEDEWGKTPDAMKAAMTLEKKLNQALLDLHALGSARTDPHLCDFLETHFLD EEVKLIKKMGDHLTNLHRLGGPEAGLGEYLFERLTLKHD		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

Table 18B. Comparison of NOV18a against NOV18b and NOV18c.

Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	1..73 35..108	49/74 (66%) 58/74 (78%)
NOV18c	1..131 1..134	81/134 (60%) 97/134 (71%)

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a

PSort analysis:	0.4500 probability located in cytoplasm; 0.3962 probability located in microbody (peroxisome); 0.1832 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV18a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. GENESEQ Results for NOV18a

GENESEQ Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP51378	Human MDDT SEQ ID NO 400 - Homo sapiens, 199 aa. [WO200240715-A2, 23-MAY-2002]	1..131 17..150	87/134 (64%) 100/134 (73%)	3e-38
AAU29923	Novel human secreted protein #414 - Homo sapiens, 238 aa. [WO200179449-A2, 25-OCT-2001]	1..131 64..197	81/134 (60%) 98/134 (72%)	7e-35
ABG16661	Novel human diagnostic protein #16652 - Homo sapiens, 313 aa. [WO200175067-A2, 11-OCT-2001]	1..131 159..292	80/134 (59%) 96/134 (70%)	3e-34
AAU29888	Novel human secreted protein #379 - Homo sapiens, 261 aa. [WO200179449-A2, 25-OCT-2001]	1..131 86..220	82/135 (60%) 98/135 (71%)	1e-33
AAU33122	Novel human secreted protein #3613 - Homo sapiens, 177 aa. [WO200179449-A2, 25-OCT-2001]	7..131 8..135	77/128 (60%) 94/128 (73%)	1e-32

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

Table 18E. Public BLASTP Results for NOV18a

Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
S01239	ferritin light chain - rabbit, 175 aa.	1..131 1..134	87/134 (64%) 99/134 (72%)	3e-38
Q9JKP6	Ferritin - Cavia porcellus (Guinea pig), 175 aa.	1..131 1..134	87/134 (64%) 101/134 (74%)	6e-38
FRRTL	ferritin light chain - rat, 183 aa.	1..131 1..134	87/134 (64%) 100/134 (73%)	8e-38
I54774	ferritin light chain - rat, 183 aa.	1..131 1..134	87/134 (64%) 100/134 (73%)	1e-37
P09451	Ferritin light chain (Ferritin L subunit) - Oryctolagus cuniculus (Rabbit), 174 aa.	2..131 1..133	86/133 (64%) 98/133 (73%)	1e-37

PFam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	**NOV18a Match Region**	**Identities/ Similarities for the Matched Region**	**Expect Value**
ferritin	13..131	54/126 (43%) 88/126 (70%)	7e-35

Example 19.

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis		
	SEQ ID NO: 57	4721 bp
NOV19a, CG56620-01 DNA Sequence	AAGTTTTTAATGATACCTGCCGCTCAGGTGGCCTAGGTGGTAGTCATGCCTTGCTTCTACTTACGTAG TTGTGGTTCTCTCTTGCCTGAACTAAAGCTTGAAGAGAGAACAGAATTTGCTCATAGGATATGGGACA CACTTCAGAAATTAGGTGCTGTGTATGATGTGAGTCACTATAATGCTTTACTTAAAGTCTATCTTCAA AATGAATATAAATTCTCACCAACTGATTTCCTGGCAAAAATGGAGGAAGCAAACATTCAACCAAATCG AGTGACATACCAGAGATTGATTGCTTCTTATTGTAATGTAGGAGATATTGAAGGTGCCAGCAAGATTC TTGGATTTATGAAAACTAAGGATCTCCCAGTTACAGAGGCAGTATTCAGTGCCCTTGTGACAGGGCAT GCCAGAGCTGGTGATATGGAGAATGCAGAAAACATTCTCACAGTGATGAGAGATGCCGGAATTGAGCC TGGTCCAGACACATACCTCGCATTATTGAATGCATATGCTGAGAAGGGCGACATTGACCATGTTAAGC AGACTCTGGAGAAGGTGGAGAAGTTCGAGCTTCACCTTATGGACCGTGATTTACTGCAAATTATTTTT AGCTTCAGTAAAGCTGGGTATCTCAGTATGTCTCAGAAATTTTGGAAAAAGTTTACATGTGAAAGAAG ATATATTCCAGATGCAATGAACCTCATTTTACTTTTAGTCACTGAAAAATTGGAAGATGTAGCGTTGC AAATTTTACTAGCATGCCCCGTATCAAAGGAAGATGGCCCAAGTGTCTTTGGCAGTTTCTTTTTACAA CACTGTGTGACTATGAATACGCCTGTGGAGAAGCTAACAGACTACTGTAAGAAGTTAAAGGAAGTCCA GATGCACTCCTTTCCTCTGCAGTTCACCCTCCATTGTGCTTTACTCGCCAATAAAACTGATTTGGCAA AAGCCTTAATGAAGGCTGTGAAGGAGGAAGGTTTTCCTATCAGACCTCACTATTTCTGGCCATTGCTA GTTGGACGTCGGAAGGAAAAAAATGTTCAAGGTATAATTGAAATCCTCAAAGGAATGCAAGAATTGGG AGTACATCCTGATCAGGAAACATATACAGATTATGTGATTCCATGCTTTGATAGTGTAAACTCAGCAC GAGCCATTTTGCAGGAAAATGGATGTCTGTCTGATAGTGATATGTTTTCTCAAGCTGGATTGAGAAGT GAAGCAGCAAATGGGAACTTAGACTTTGTATTATCATTTTTGAAATCAAATACATTGCCCATCTCGCT GCAGTCTATAAGAAGTAGCCTACTGCTAGGCTTCAGGAGGTCTATGAATATAAATGTTTGGAGCGAGA TAACAGAATTATTGTACAAGGATGGACGTTATTGCCAGGAGCCTCGAGGACCGACGGAAGCTGTTGGC AATTTTCTTTATAACTTGATTGACAGCATGAGTGACTCAGAGGTACAGGCCAAGGAGGAGCATTTGAG ACAATACTTCCATCAGCTGGAGAAGATGAATGTAAAAATTCCTGAAAATATCTACAGAGGCATTCGTA ATCTCCTGGAAAGCTACCATGTTCCTGAATTGATTAAGGATGCTCACTTGTTGGTTGAGCGTAAGAAT TTAGACTTTCAAAAAACTGTGCAACTTACATCATCTGAATTGGAGTCAACACTTGAAACACTAAAAGC TGAAAATCAACCTATAAGAGATGTCCTAAAGCAACTCATATTAGTGCTTTGTTCAGAAGAGAATATGC AAAAAGCCCTTGAATTGAAAGCAAAATATGAATCCGACATGGTTACTGGTGGCTATGCAGCTTTAATA AATTTATGCTGTCGACATGATAAAGTAGAAGATGCCTTGAACTTGAAAGAAGAATTTGACCGCTTAGA TTCATCTGCTGTCCTTGACACCGGCAACTATCTAGGCCTTGTAAGAGTATTGGCAAAGCATGGCAAGC TCCAAGATGCTATTAAGATTCTGAAGGAGATGAAAGAGAAGGATGTTCTTATCAAAGATACAACAGCC TTGTCCTTTTTCCACATGCTAAATGGCGCAGCTTTAAGAGGTGAAATTGAAACAGTAAAACAGTTGCA TGAAGCCATCGTGACTCTAGGGTTAGCAGAACCATCCACCAACATAAGTTTCCCATTGGTCACTGTAC ACTTGGAAAAGGGCGACCTATCTACTGCTCTTGAGGTCGCCATTGACTGCTATGAAAAGTATAAAGTA TTACCAAGGATTCATGATGTCTTGTGTAAACTGGTAGAGAAAGGCGAGACTGATCTAATTCAGAAAGC AATGGACTTTGTGAGCCAAGAACAAGGTGAAATGGTGATGCTCTATGATCTCTTCTTTGCCTTCCTAC AAACAGGAAATTACAAAGAGGCCAAGAAGATCATTGAGACTCCAGGGATTAGAGCTCGATCTGCAAGG CTTCAGTGGTTTTGTGACAGATGTGTTGCAAATAATCAGGTTGAAACTCTGGAAAAATTAGTGGAGCT GACACAGAAGCTATTTGAATGTGATAGAGACCAGATGTACTACAATCTGCTAAAACTGTATAAAATAA ACGGTGACTGGCAAAGAGCTGATGCAGTCTGGAATAAAATCCAAGAAGAAAATGTTATTCCTCGTGAA AAGACATTAAGATTATTAGCAGAAATCCTTAGAGAGGGTAACCAGGAAGTTCCGTTTGACGTACCTGA GTTGTGGTATGAAGATGAAAAACATTCCCTGAATTCTTCGTCAGCCTCAACCACAGAACCTGATTTCC AGAAAGATATATTGATTGCCTGCCGATTGAACCAAAAAAAAGGGGCATATGATATTTTCCTGAATGCA AAAGAGCAAAACATTGTGTTTAATGCTGAAACCTACAGCAATCTCATTAAATTACTGATGTCAGAAGA TTATTTTACACAAGCAATGGAAGTGAAAGCATTCGCGGAGACCCACATCAAGGGCTTCACACTGAACG

	ATGCTGCCAACAGCCGCCTCATCATAACGCAAGTTAGGCGGGATTATTTGAAAGAGGCTGTGACAACA CTGAAAACAGTATTGGATCAGCAGCAGACCCCTTCTAGGTTAGCAGTGACCCGTGTCATCCAGGCATT GGCCATGAAGGGTGATGTTGAAAACATAGAAGTAGTTCAGAAGATGTTAAATGGACTCGAAGACTCCA TTGGACTTTCAAAAATGGTTTTCATCAATAACATTGCTTTGGCTCAAATAAAGAATAATAACATAGAT GCCGCAATAGAAAACATTGAAAATATGCTTACTTCAGAGAATAAAGTCATTGAACCCCAATACTTCGG CTTGGCATACTTATTCAGAAAAGTAATAGAGGAGCAGTTGGAACCAGCAGTTGAAAAGATAAGCATCA TGGCGGAGAGATTGGCCAATCAGTTTGCAATTTATAAACCTGTCACTGATTTTTTCCTTCAACTTGTG GATGCAGGCAAGGTGGATGATGCCAGAGCTCTCCTACAGAGATGTGGTGCAATTGCTGAACAAACCCC GATTTTGTTGTTGTTCCTCCTTAGGAATTCTAGGAAACAAGGAAAGGCATCAACTGTGAAATCTGTGT TAGAATTGATTCCTGAATTAAATGAAAAGGAAGAAGCATACAATTCCCTCATGAAAAGCTATGTCTCA GAGAAAGATGTCACATCTGCTAAAGCACTGTATGAACATTTGACTGCAAAGAATACAAAATTGGATGA TCTGTTTCTAAAGCGTTACGCATCTTTGCTGAAGTATGCTGGAGAGCCTGTCCCTTTCATTGAACCCC CTGAAAGCTTTGAATTTTATGCACAGCAGCTAAGAAAATTGAGGGAAAACTCTTCTTGAAATAACCAG GCGATACTTTGTTTTGTATATATTTGTGATTCTGTGTCTACATGTTATTTTGAAGTATATCTGAGGGA AAAATAAATGAAAATTTTCTTTATGTACTTATGTATGTGTGATGCATGTTCAAAGTCTTATTGACCAT AACTCTGTGCACTTGGTTATTGGACATTTTTGGAGTTTTTTCTCTGGGAAAAATCGATAGTGTTTTCT TCAATGCTGCTGCTGTGTGAAGCCATACTTTTCAGGATTCTTCCCTAATTGGCTCTTTGGTTTCCCTG CTCTGTTTCATTTATTTCATTAAAATGTTATTCCTTTATTTAAGATTCACTTATTAGTCTGCTGTTTC TCTGAAAAATTTTAGAGCTAGGTATAGTGACCGTGAACTTCTAACGCATAATATCTGTGATACAGCCA TTCCGTACATGTGTGAGTCTGCATAACTTTCGAACTTTGTTAAATGTTGGCACTAGGAGT CATCAGATCTAGGATTCATCATTTTCCAGTGAGAAGCAGAGACCCAAAGGCCTGTTACTTGTGCTTGG TCAGGGGACTGTCTGTCATGCCTGGAGGCTCTTCGGCACACTTCCCCATCTTTCCCTTCTGCCACTGT GGCTTCAAGCACCTCTGTTCATAGAGCGTCTCTGAAATTGAGTCTCGGTCATGACTTATCCCGAAGTA GAGCAATGTGTTTCCTCTCATTGTAGTTTCAGGACTTTGTCAGTACAAGCTCTGCCCTAGGCTTCTTA CTTTATACTCATATCCTGAAAAGATGTGATTTCATCTATGAAGGGGTAAAATATTGGTTTGTATTTAA TTGTTTGAAATAAAAGTGATCCCTATAAA
	ORF Start: ATG at 46 | | ORF Stop: TGA at 3865
	SEQ ID NO: 58 | 1273 aa | MW at 145199.5kD
NOV19a, CG56620-01 Protein Sequence	MPCFYLRSCGSLLPELKLEERTEFAHRIWDTLQKLGAVYDVSHYNALLKVYLQNEYKFSPTDFLAKME EANIQPNRVTYQRLIASYCNVGDIEGASKILGFMKTKDLPVTEAVFSALVTGHARAGDMENAENILTV MRDAGIEPGPDTYLALLNAYAEKGDIDHVKQTLEKVEKFELHLMDRDLLQIIFSFSKAGYLSMSQKFW KKFTCERRYIPDAMNLILLLVTEKLEDVALQILLACPVSKEDGPSVFGSFFLQHCVTMNTPVEKLTDY CKKLKEVQMHSFPLQFTLHCALLANKTDLAKALMKAVKEEGFPIRPHYFWPLLVGRRKEKNVQGIIEI LKGMQELGVHPDQETYTDYVIPCFDSVNSARAILQENGCLSDSDMFSQAGLRSEAANGNLDFVLSFLK SNTLPISLQSIRSSLLLGFRRSMNINVWSEITELLYKDGRYCQEPRGPTEAVGNFLYNLIDSMSDSEV QAKEEHLRQYFHQLEKMNVKIPENIYRGIRNLLESYHVPELIKDAHLLVERKNLDFQKTVQLTSSELE STLETLKAENQPIRDVLKQLILVLCSEENMQKALELKAKYESDMVTGGYAALINLCCRHDKVEDALNL KEEFDRLDSSAVLDTGNYLGLVRVLAKHGKLQDAIKILKEMKEKDVLIKDTTALSFFHMLNGAALRGE IETVKQLHEAIVTLGLAEPSTNISFPLVTVHLEKGDLSTALEVAIDCYEKYKVLPRIHDVLCKLVEKG ETDLIQKAMDFVSQEQGEMVMLYDLFFAFLQTGNYKEAKKIIETPGIRARSARLQWFCDRCVANNQVE TLEKLVELTQKLFECDRDQMYYNLLKLYKINGDWQRADAVWNKIQEENVIPREKTLRLLAEILREGNQ EVPFDVPELWYEDEKHSLNSSSASTTEPDFQKDILIACRLNQKKGAYDIFLNAKEQNIVFNAETYSNL IKLLMSEDYFTQAMEVKAFAETHIKGFTLNDAANSRLIITQVRRDYLKEAVTTLKTVLDQQQTPSRLA VTRVIQALAMKGDVENIEVVQKMLNGLEDSIGLSKMVFINNIALAQIKNNNIDAAIENIENMLTSENK VIEPQYFGLAYLFRKVIEEQLEPAVEKISIMAERLANQFAIYKPVTDFFLQLVDAGKVDDARALLQRC GAIAEQTPILLLFLLRNSRKQGKASTVKSVLELIPELNEKEEAYNSLMKSYVSEKDVTSAKALYEHLT AKNTKLDDLFLKRYASLLKYAGEPVPFIEPPESFEFYAQQLRKLRENSS
	SEQ ID NO: 59 | 3757 bp |
NOV19b, CG56620-02 DNA Sequence	AAGTTTTTAATGATACCTGCCGCTCAGGTGGCCTAGGTGGTAGTCATGCCTTGCTTCTACTTACGTAG TTGTGGTTCTCTCTTGCCTGAACTAAAGCTTGAAGAGAGAACAGAATTTGCTCATAGGATATGGGACA CACTTCAGAAATTAGGTGCTGTGTATGATGTGAGTCACTATAATGCTTTACTTAAAGTCTATCTTCAA AATGAATATAAATTCTCACCAACTGATTTCCTGGCAAAAATGGAGGAAGCAAACATTCAACCAAATCG AGTGACATACCAGAGATTGATTGCTTCTTATTGTAATGTAGGAGATATTGAAGGTGCCAGCAAGATTC TTGGATTTATGAAAACTAAGGATCTCCCAGTTACAGAGGCAGTATTCAGTGCCCTTGTGACAGGGCAT GCCAGAGCTGGTGATATGGAGAATGCAGAAAACATTCTCACAGTGATGAGAGATGCCGGAATTGAGCC TGGTCCAGACACATACCTCGCATTATTGAATGCATATGCTGAGAAGGGCGACATTGACCATGTTAAGC AGACTCTGGAGAAGGTGGAGAAGTTCGAGCTTCACCTTATGGACCGTGATTTACTGCAAATTATTTTT AGCTTCAGTAAAGCTGGGTATCTCAGTATGTCTCAGAAATTTTGGAAAAAGTTTACATGTGAAAGAAG ATATATTCCAGATGCAATGAACCTCATTTTACTTTTAGTCACTGAAAAATTGGAAGATGTAGCGTTGC AAATTTTACTAGCATGCCCCGTATCAAAGGAAGATGGCCCAAGTGTCTTTGGCAGTTTCTTTTTACAA CACTGTGTGACTATGAATACGCCTGTGGAGAAGCTAACAGACTACTGTAAGAAGTTAAAGGAAGTCCA GATGCACTCCTTTCCTCTGCAGTTCACCCTCCATTGTGCTTTACTCGCCAATAAAACTGATTTGGCAA AAGCCTTAATGAAGGCTGTGAAGGAGGAAGGTTTTCCTATCAGACCTCACTATTTCTGGCCATTGCTA GTTGGACGTCGGAAGGAAAAAAATGTTCAAGGTATAATTGAAATCCTCAAAGGAATGCAAGAATTGGG AGTACATCCTGATCAGGAAACATATACAGATTATGTGATTCCATGCTTTGATAGTGTAAACTCAGCAC

	GAGCCATTTTGCAGGAAAATGGATGTCTGTCTGATAGTGATATGTTTTCTCAAGCTGGATTGAGAAGT GAAGCAGCAAATGGGAACTTAGACTTTGTATTATCATTTTTGAAATCAAATACATTGCCCATCTCGCT GCAGTCTATAAGAAGTAGCCTACTGCTAGGCTTCAGGAGGTCTATGAATATAAATGTTTGGAGCGAGA TAACAGAATTATTGTACAAGGATGGACGTTATTGCCAGGAGCCTCGAGGACCGACGGAAGCTGTTGGC AATTTTCTTTATAACTTGATTGACAGCATGAGTGACTCAGAGGTACAGGCCAAGGAGGAGCATTTGAG ACAATACTTCCATCAGCTGGAGAAGATGAATGTAAAAATTCCTGAAAATATCTACAGAGGCATTCGTA ATCTCCTGGAAAGCTACCATGTTCCTGAATTGATTAAGGATGCTCACTTGTTGGTTGAGCGTAAGAAT TTAGACTTTCAAAAAACTGTGCAACTTACATCATCTGAATTGGAGTCAACACTTGAAACACTAAAAGC TGAAAATCAACCTATAAGAGATGTCCTAAAGCAACTCATATTAGTGCTTTGTTCAGAAGAGAATATGC AAAAAGCCCTTGAATTGAAAGCAAAATATGAATCCGACATGGTTACTGGTGGCTATGCAGCTTTAATA AATTTATGCTGTCGACATGATAAAGTAGAAGATGCCTTGAACTTGAAAGAAGAATTTGACCGCTTAGA TTCATCTGCTGTCCTTGACACCGGCAACTATCTAGGCCTTGTAAGAGTATTGGCAAAGCATGGCAAGC TCCAAGATGCTATTAAGATTCTGAAGGAGATGAAAGAGAAGGATGTTCTTATCAAAGATACAACAGCC TTGTCCTTTTTCCACATGCTAAATGGCGCAGCTTTAAGAGGTGAAATTGAAACAGTAAAACAGTTGCA TGAAGCCATCGTGACTCTAGGGTTAGCAGAACCATCCACCAACATAAGTTTCCCATTGGTCACTGTAC ACTTGGAAAAGGGCGACCTATCTACTGCTCTTGAGGTCGCCATTGACTGCTATGAAAAGTATAAAGTA TTACCAAGGATTCATGATGTCTTGTGTAAACTGGTAGAGAAAGGCGAGACTGATCTAATTCAGAAAGC AATGGACTTTGTGAGCCAAGAACAAGGTGAAATGGTGATGCTCTATGATCTCTTCTTTGCCTTCCTAC AAACAGGAAATTACAAAGAGGCCAAGAAGATCATTGAGACTCCAGGGATTAGAGCTCGATCTGCAAGG CTTCAGTGGTTTTGTGACAGATGTGTTGCAAATAATCAGGTTGAAACTCTGGAAAAATTAGTGGAGCT GACACAGAAGCTATTTGAATGTGATAGAGACCAGATGTACTACAATCTGCTAAAACTGTATAAAATAA ACGGTGACTGGCAAAGAGCTGATGCAGTCTGGAATAAAATCCAAGAAGAAAATGTTATTCCTCGTGAA AAGACATTAAGATTATTAGCAGAAATCCTTAGAGAGGGTAACCAGGAAGTTCCGTTTGACGTACCTGA GTTGTGGTATGAAGATGAAAAACATTCCCTGAATTCTTCGTCAGCCTCAACCACAGAACCTGATTTCC AGAAAGATATATTGATTGCCTGCCGATTGAACCAAAAAAAAGGGGCATATGATATTTTCCTGAATGCA AAAGAGCAAAACATTGTGTTTAATGCTGAAACCTACAGCAATCTCATTAAATTACTGATGTCAGAAGA TTATTTTACACAAGCAATGGAAGTGAAAGCATTCGCGGAGACCCACATCAAGGGCTTCACACTGAACG ATGCTGCCAACAGCCGCCTCATCATAACGCAAGTTAGGCGGGATTATTTGAAAGAGGCTGTGACAACA CTGAAAACAGTATTGGATCAGCAGCAGACCCCTTCTAGGTTAGCAGTGACCCGTGTCATCCAGGCATT GGCCATGAAGGGTGATGTTGAAAACATAGAAGTAGTTCAGAAGATGTTAAATGGACTCGAAGACTCCA TTGGACTTTCAAAAATGGTTTTCATCAATAACATTGCTTTGGCTCAAATAAAGAATAATAACATAGAT GCCGCAATAGAAAACATTGAAAATATGCTTACTTCAGAGAATAAAGTCATTGAACCCCAATACTTCGG CTTGGCATACTTATTCAGAAAAGTAATAGAGGAGCAGTTGGAACCAGCAGTTGAAAAGATAAGCATCA TGGCGGAGAGATTGGCCAATCAGTTTGCAATTTATAAACCTGTCACTGATTTTTTCCTTCAACTTGTG GATGCAGGCAAGGTGGATGATGCCAGAGCTCTCCTACAGAGATGTGGTGCAATTGCTGAACAAACCCC GATTTTGTTGTTGTTCCTCCTTAGGAATTCTAGGAAACAAGGAAAGGCATCAACTGTGAAATCTGTGT TAGAATTGATTCCTGAATTAAATGAAAAGGAAGAAGCATACAATTCCCTCATGAAAAATTTTAGAGCT AGGTATAGTGACCGTGAACTTTCTAACGCATAATATTCTGTGATACAGCCATTCCGTACATGTGTGAA GTCCTGCATAACTTTCG		
	ORF Start: ATG at 46		ORF Stop: TAA at 3703
	SEQ ID NO: 60	1219 aa	MW at 139025.5kD
NOV19b, CG56620-02 Protein Sequence	MPCFYLRSCGSLLPELKLEERTEFAHRIWDTLQKLGAVYDVSHYNALLKVYLQNEYKFSPTDFLAKME EANIQPNRVTYQRLIASYCNVGDIEGASKILGFMKTKDLPVTEAVFSALVTGHARAGDMENAENILTV MRDAGIEPGPDTYLALLNAYAEKGDIDHVKQTLEKVEKFELHLMDRDLLQIIFSFSKAGYLSMSQKFW KKFTCERRYIPDAMNLILLLVTEKLEDVALQILLACPVSKEDGPSVFGSFFLQHCVTMNTPVEKLTDY CKKLKEVQMHSFPLQFTLHCALLANKTDLAKALMKAVKEEGFPIRPHYFWPLLVGRRKEKNVQGIIEI LKGMQELGVHPDQETYTDYVIPCFDSVNSARAILQENGCLSDSDMFSQAGLRSEAANGNLDFVLSFLK SNTLPISLQSIRSSLLLGFRRSMNINVWSEITELLYKDGRYCQEPRGPTEAVGNFLYNLIDSMSDSEV QAKEEHLRQYFHQLEKMNVKIPENIYRGIRNLLESYHVPELIKDAHLLVERKNLDFQKTVQLTSSELE STLETLKAENQPIRDVLKQLILVLCSEENMQKALELKAKYESDMVTGGYAALINLCCRHDKVEDALNL KEEFDRLDSSAVLDTGNYLGLVRVLAKHGKLQDAIKILKEMKEKDVLIKDTTALSFFHMLNGAALRGE IETVKQLHEAIVTLGLAEPSTNISFPLVTVHLEKGDLSTALEVAIDCYEKYKVLPRIHDVLCKLVEKG ETDLIQKAMDFVSQEQGEMVMLYDLFFAFLQTGNYKEAKKIIETPGIRARSARLQWFCDRCVANNQVE TLEKLVELTQKLFECDRDQMYYNLLKLYKINGDWQRADAVWNKIQEENVIPREKTLRLLAEILREGNQ EVPFDVPELWYEDEKHSLNSSSASTTEPDFQKDILIACRLNQKKGAYDIFLNAKEQNIVFNAETYSNL IKLLMSEDYFTQAMEVKAFAETHIKGFTLNDAANSRLIITQVRRDYLKEAVTTLKTVLDQQQTPSRLA VTRVIQALAMKGDVENIEVVQKMLNGLEDSIGLSKMVFINNIALAQIKNNNIDAAIENIENMLTSENK VIEPQYFGLAYLFRKVIEEQLEPAVEKISIMAERLANQFAIYKPVTDFFLQLVDAGKVDDARALLQRC GAIAEQTPILLLFLLRNSRKQGKASTVKSVLELIPELNEKEEAYNSLMKNFRARYSDRELSNA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 19B.

Table 19B. Comparison of NOV19a against NOV19b.		
Protein Sequence	**NOV19a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV19b	1..1216 1..1219	1165/1219 (95%) 1173/1219 (95%)

Further analysis of the NOV19a protein yielded the following properties shown in Table 19C.

Table 19C. Protein Sequence Properties NOV19a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV19a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19D.

Table 19D. GENESEQ Results for NOV19a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV19a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU75103	Leucine rich protein, L130 - Homo sapiens, 1273 aa. [WO200198524-A2, 27-DEC-2001]	1..1273 1..1273	1273/1273 (100%) 1273/1273 (100%)	0.0
ABB71881	Drosophila melanogaster polypeptide SEQ ID NO 42435 - Drosophila melanogaster, 1275 aa. [WO200171042-A2, 27-SEP-2001]	40..1045 2..1027	314/1035 (30%) 512/1035 (49%)	e-129
AAY95854	Autoantigen diagnostic of endometriosis - Homo sapiens, 189 aa. [WO200047739-A2, 17-AUG-2000]	516..700 1..185	179/185 (96%) 180/185 (96%)	7e-95
AAO00059	Human polypeptide SEQ ID NO 13951 - Homo sapiens, 164 aa. [WO200164835-A2, 07-SEP-2001]	965..1124 1..160	137/160 (85%) 146/160 (90%)	1e-69

ABB69797	Drosophila melanogaster polypeptide SEQ ID NO 36183 - Drosophila melanogaster, 1072 aa. [WO200171042-A2, 27-SEP-2001]	4..887 125..1062	219/972 (22%) 405/972 (41%)	6e-41

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19E.

Table 19E. Public BLASTP Results for NOV19a

Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P42704	130 kDa leucine-rich protein (LRP 130) (GP130) (Leucine-rich PPR-motif containing protein) - Homo sapiens (Human), 1273 aa.	1..1273 1..1273	1273/1273 (100%) 1273/1273 (100%)	0.0
S27954	leucine-rich protein - human, 1207 aa.	67..1273 1..1207	1207/1207 (100%) 1207/1207 (100%)	0.0
BAB93528	Leucine rich protein mLRP130 - Mus musculus (Mouse), 1306 aa.	6..1273 40..1306	955/1268 (75%) 1100/1268 (86%)	0.0
Q96D84	Hypothetical 80.0 kDa protein - Homo sapiens (Human), 702 aa (fragment).	572..1273 1..702	699/702 (99%) 700/702 (99%)	0.0
Q9CRX4	3110001K13Rik protein - Mus musculus (Mouse), 712 aa (fragment).	561..1273 1..712	542/713 (76%) 627/713 (87%)	0.0

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19F.

Table 19F. Domain Analysis of NOV19a

Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PPR	77..111	13/35 (37%) 23/35 (66%)	0.0032
PPR	112..146	9/35 (26%) 25/35 (71%)	0.00018
PPR	628..662	12/35 (34%) 27/35 (77%)	0.046

PPR	1198..1232	10/35 (29%) 24/35 (69%)	0.037

Example 20.

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

5

Table 20A. NOV20 Sequence Analysis

	SEQ ID NO: 61	4534 bp
NOV20a, CG59323-01 DNA Sequence	GTCACGAGCGTCGAAGAGACAAAGCCGCGTCAGGGGGCCCGGCCGGGGCGGGGGAGCCCGGGGCTTGTTG GTGCCCCAGCCCGCGCGGAGGGCCCTTCGGACCCGCGCGCCGCCGCTGCCGCCGCCGCCGCCTCGCAACA GGTCCGGGCGGCCTCGCTCTCCGCTCCCCTCCCCCGCATCCGCGACCCTCCGGGGCACCTCAGCTCGGCC GGGGCCGCAGTCTGGCCACCCGCTTCCATGCGGTTCGGGTCCAAGATGATGCCGATGTTTCTTACCGTGT ATCTCAGTAACAATGAGCAGCACTTCACAGAAGTTCCAGTTACTCCAGAAACAATATGCAGAGACGTGGT GGATCTGTGCAAAGAACCCGGCGAGAGTGATTGCCATTTGGCTGAAGTGTGGTGTGGCTCTGTAGAGATA GAGTTTCATCATGTTGGCCAGGATGGTCTCGATCTCCTGACCTTGTGATCCGCCTGCCTCGGCCTCCCAA AGTGCTGGATTACAGGTGTGAGCCACCACGATCAGCCTCTAGTGTTTAAAAAAGAACGTCCAGTTGCGGA TAATGAGCGAATGTTTGATGTTCTTCAACGATTTGGAAGTCAGAGGAACGAAGTTCGCTTCTTCCTTCGT CATGAACGCCCCCCTGGCAGGGACATTGTGAGTGGACCAAGATCTCAGGATCCAAGTTTAAAAAGAAATG GTGTAAAAGTTCCTGGTGAATATCGAAGAAAGGAGAACGGTGTTAATAGTCCTAGGATGGATCTGACTCT TGCTGAACTTCAGGAAATGGCATCTCGCCAGCAGCAACAGATTGAAGCCCAGCAACAATTGCTGGCAACT AAGGAACAGCGCTTAAAGTTTTTGAAACAACAAGATCAGCGACAACAGCAACAAGTTGCTGAGCAGGAGA AACTTAAAAGGCTAAAAGAAATAGCTGAGAATCAGGAAGCTAAGCTAAAAAAAGTGAGAGCACTTAAAGG CCACGTGGAACAGAAGAGACTAAGCAATGGGAAACTTGTGGAGGAAATTGAACAGATGAATAATTTGTTC CAGCAAAAACAGAGGGAGCTCGTCCTGGCTGTGTCAAAAGTAGAAGAACTGACCAGGCAGCTAGAGATGC TCAAGAACGGCAGGATCGACAGCCACCATGACAATCAGTCTGCAGTGGCTGAGCTTGATCGCCTCTATAA GGAGCTGCAGCTAAGAAACAAATTGAATCAAGAGCAGAATGCCAAGCTACAACAACAGAGGGAGTGTTTG AATAAGCGTAATTCAGAAGTGGCAGTCATGGATAAGCGTGTTAATGAGCTGAGGGACCGGCTGTGGAAGA AGAAGGCAGCTCTACAGCAAAAAGAAAATCTACCAGTTTCATCTGATGGAAATCTTCCCCAGCAAGCCGC GTCAGCCCCAAGCCGTGTGGCTGCAGTAGGTCCCTATATCCAGTCATCTACTATGCCTCGGATGCCCTCA AGGCCTGAATTGCTGGTGAAGCCAGCCCTGCCGGATGGTTCCTTGGTCATTCAGGCTTCAGAGGGGCCGA TGAAAATACAGACACTGCCCAACATGAGATCTGGGGCTGCTTCACAAACTAAAGGCTCTAAAATCCATCC AGTTGGCCCTGATTGGAGTCCTTCAAATGCAGATCTTTTCCCAAGCCAAGGCTCTGCTTCTGTACCTCAA AGCACTGGGAATGCTCTGGATCAAGTTGATGATGGAGAGGTTCCGCTGAGGGAGAAAGAGAAGAAAGTGC GTCCGTTCTCAATGTTTGATGCAGTAGACCAGTCCAATGCCCCACCTTCCTTTGGTACTCTGAGGAAGAA CCAGAGCAGTGAAGATATCTTGCGGGATGCTCAGGTTGCAAATAAAAATGTGGCTAAAGTACCACCTCCT GTTCCTACAAAACCAAAACAGATTAATTTGCCTTATTTTGGACAAACTAATCAGCCACCTTCAGACATTA AGCCAGACGGAAGTTCTCAGCAGTTGTCAACAGTTGTTCCGTCCATGGGAACTAAACCAAAACCAGCAGG GCAGCAGCCGAGAGTGCTGCTATCTCCCAGCATACCTTCGGTTGGCCAAGACCAGACCCTTTCTCCAGGT TCTAAGCAAGAAAGTCCACCTGCTGCTGCCGTCCGGCCCTTTACTCCCCAGCCTTCCAAAGACACCTTAC TTCCACCCTTCAGAAAACCCCAGACCGTGGCAGCAAGTTCAATATATTCCATGTATACGCAACAGCAGGC GCCAGGAAAAAACTTCCAGCAGGCTGTGCAGAGCGCGTTGACCAAGACTCATACCAGAGGGCCACACTTT TCAAGTGTATATGGTAAGCCTGTAATTGCTGCTGCCCAGAATCAACAGCAGCACCCAGAGAACATTTATT CCAATAGCCAGGGCAAGCCTGGCAGTCCAGAACCTGAAACAGAGCCTGTTTCTTCAGTTCAGGAGAACCA TGAAAACGAAAGAATTCCTCGGCCACTCAGCCCAACTAAATTACTGCCTTTCTTATCTAATCCTTACCGA AACCAGAGTGATGCTGACCTAGAAGCCTTACGAAAGAAACTGTCTAACGCACCAAGGCCTCTAAAGAAAC GTAGTTCTATTACAGAGCCAGAGGGTCCTAATGGGCCAAATATTCAGAAGCTTTTATATCAGAGGACCAC CATAGCGGCCATGGAGACCATCTCTGTCCCATCATACCCATCCAAGTCAGCTTCTGTGACTGCCAGCTCA GAAAGCCCAGTAGAAATCCAGAATCCATATTTACATGTGGAGCCCGAAAAGGAGGTGGTCTCTCTGGTTC CTGAATCATTGTCCCCAGAGGATGTGGGGAATGCCAGTACAGAGAACAGTGACATGCCAGCTCCTTCTCC AGGCCTTGATTATGAGCCTGAGGGAGTCCCAGACAACAGCCCAAATCTCCAGAATAACCCAGAAGAACCA AATCCAGAGGCTCCACATGTGCTTGATGTGTACCTGGAGGAGTACCCTCCATACCCACCCCCACCATACC CATCTGGGGAGCCTGAAGGGCCCGGAGAAGACTCGGTGAGCATGCGCCCGCCTGAAATCACCGGGCAGGT CTCTCTGCCTCCTGGTAAAAGGACAAACTTGCGTAAAACTGGCTCAGAGCGTATCGCTCATGGAATGAGG GTGAAATTCAACCCCCTTGCTTTACTGCTAGATTCGTCTTTGGAGGGAGAATTTGACCTTGTACAGAGAA TTATTTATGAGGTTGATGACCCAAGCCTGCCCAATGATGAAGGCATCACGGCTCTTCACAATGCTGTGTG TGCAGGCCACACAGAAATCGTTAAGTTCCTGGTACAGTTTGGTGTAAATGTAAATGCTGCTGATAGTGAT GGATGGACTCCATTACATTGTGCTGCCTCATGTAACAACGTCCAAGTGTGTAAGTTTTTGGTGGAGTCAG GAGCCGCTGTGTTTGCCATGACCTACAGTGACATGCAGACTGCTGCAGATAAGTGCGAGGAAATGGAGGA AGGCTACACTCAGTGCTCCCAATTTCTTTATGGAGTTCAGGAGAAGATGGGCATAATGAATAAAGGAGTC	

	ATTTATGCGCTTTGGGATTATGAACCTCAGAATGATGATGAGCTGCCCATGAAAGAAGGAGACTGCATGA CAATCATCCACAGGGAAGACGAAGATGAAATCGAATGGTGGTGGGCGCGCCTTAATGATAAGGAGGGATA TGTTCCACGTAACTTGCTGGGACTGTACCCAAGAATTAAACCAAGACAAAGGAGCTTGGCCTGAAACTTC CACACAGAATTTTAGTCAATGAAGAATTAATCTCTGTTAAGAAGAAGTAATACGATTATTTTTGGCAAAA ATTTCACAAGACTTATTTTAATGACAATGTAGCTTGAAAGCGATGAAGAATGTCTCTAGAAGAGAATGAA GGATTGAAGAATTCACCATTAGAGGACATTTAGCGTGATGAAATAAAGCATCTACGTCAGCAGGCCATAC TGTGTTGGGGCAAAGGTGTCCCGTGTAGCACTCAGATAAGTATACAGCGACAATCCTGTTTTCTACAAGA ATCCTGTCTAGTAAATAGGATCATTTATTGGGCAGTTGGGAAATCAGCTCTCTGTCCTGTTGAGTGTTTT CAGCAGCTGCTCCTAAACCAGTCCTCCTGCCAGAAAGGACCAGTGCCGTCACATCGCTGTCTCTGATTGT CCCCGGCACCAGCAGGCCTTGGGGCTCACTGAAGGCTCGAAGGCACTGCACACCTTGTATATTGTCAGTG AAGAACGTTAGTTGGTTGTCAGTGAACAATAACTTTATTATATGAGTTTTTGTAGCATCTTAAGAATTAT ACATATGTTTGAAATATTGAAACTAAGCTACAGTACCAGTAATTAGATGTAGAATCTTGTTTGTAGGCTG AATTTTAATCTGTATTTATTGTCTTTTGTATCTCAGAAATTAGAAACTTGCTACAGACTTACCCGTAATA TTTGTCAAGATCATAGCTGACTTTAAAAACAGTTGTAATAAACTTTTTGATGCT
	ORF Start: ATG at 757 | | ORF Stop: TGA at 3772
	SEQ ID NO: 62 | 1005 aa | MW at 111430.4kD
NOV20a, CG59323-01 Protein Sequence	MDLTLAELQEMASRQQQQIEAQQQLLATKEQRLKFLKQQDQRQQQQVAEQEKLKRLKEIAENQEAKLKKV RALKGHVEQKRLSNGKLVEEIEQMNNLFQQKQRELVLAVSKVEELTRQLEMLKNGRIDSHHDNQSAVAEL DRLYKELQLRNKLNQEQNAKLQQQRECLNKRNSEVAVMDKRVNELRDRLWKKKAALQQKENLPVSSDGNL PQQAASAPSRVAAVGPYIQSSTMPRMPSRPELLVKPALPDGSLVIQASEGPMKIQTLPNMRSGAASQTKG SKIHPVGPDWSPSNADLFPSQGSASVPQSTGNALDQVDDGEVPLREKEKKVRPFSMFDAVDQSNAPPSFG TLRKNQSSEDILRDAQVANKNVAKVPPPVPTKPKQINLPYFGQTNQPPSDIKPDGSSQQLSTVVPSMGTK PKPAGQQPRVLLSPSIPSVGQDQTLSPGSKQESPPAAAVRPFTPQPSKDTLLPPFRKPQTVAASSIYSMY TQQQAPGKNFQQAVQSALTKTHTRGPHFSSVYGKPVIAAAQNQQQHPENIYSNSQGKPGSPEPETEPVSS VQENHENERIPRPLSPTKLLPFLSNPYRNQSDADLEALRKKLSNAPRPLKKRSSITEPEGPNGPNIQKLL YQRTTIAAMETISVPSYPSKSASVTASSESPVEIQNPYLHVEPEKEVVSLVPESLSPEDVGNASTENSDM PAPSPGLDYEPEGVPDNSPNLQNNPEEPNPEAPHVLDVYLEEYPPYPPPPYPSGEPEGPGEDSVSMRPPE ITGQVSLPPGKRTNLRKTGSERIAHGMRVKFNPLALLLDSSLEGEFDLVQRIIYEVDDPSLPNDEGITAL HNAVCAGHTEIVKFLVQFGVNVNAADSDGWTPLHCAASCNNVQVCKFLVESGAAVFAMTYSDMQTAADKC EEMEEGYTQCSQFLYGVQEKMGIMNKGVIYALWDYEPQNDDELPMKEGDCMTIIHREDEDEIEWWWARLN DKEGYVPRNLLGLYPRIKPRQRSLA
	SEQ ID NO: 63 | 4204 bp |
NOV20b, CG59323-03 DNA Sequence	GTCACGAGCGTCGAAGAGACAAAGCCGCGTCAGGGGGCCCGGCCGGGGCGGGGGAGCCCGGGGCTTGTTG GTGCCCCAGCCCGCGCGGAGGGCCCTTCGGACCCGCGCGCCGCCGCTGCCGCCGCCGCCGCCTCGCAACA GGTCCGGGCGGCCTCGCTCTCCGCTCCCCTCCCCCGCATCCGCGACCCTCCGGGGCACCTCAGCTCGGCC GGGGCCGCAGTCTGGCCACCCGCTTCCATGCGGTTCGGGTCCAAGATGATGCCGATGTTTCTTACCGTGT ATCTCAGTAACAATGAGCAGCACTTCACAGAAGTTCCAGTTACTCCAGAAACAATATGCAGAGACGTGGT GGATCTGTGCAAAGAACCCGGCGAGAGTGATTGCCATTTGGCTGAAGTGTGGTGTGGCTCTGAACGTCCA GTTGCGGATAATGAGCGAATGTTTGATGTTCTTCAACGATTTGGAAGTCAGAGGAACGAAGTTCGCTTCT TCCTTCGTCATGAACGCCCCCCTGGCAGGGACATTGTGAGTGGACCAAGATCTCAGGATCCAAGTTTAAA AAGAAATGGTGTAAAAGTTCCTGGTGAATATCGAAGAAAGGAGAACGGTGTTAATAGTCCTAGGATGGAT CTGACTCTTGCTGAACTTCAGGAAATGGCATCTCGCCAGCAGCAACAGATTGAAGCCCAGCAACAATTGC TGGCAACTAAGGAACAGCGCTTAAAGTTTTTGAAACAACAAGATCAGCGACAACAGCAACAAGTTGCTGA GCAGGAGAAACTTAAAAGGCTAAAAGAAATAGCTGAGAATCAGGAAGCTAAGCTAAAAAAAGTGAGAGCA CTTAAAGGCCACGTGGAACAGAAGAGACTAAGCAATGGGAAACTTGTGGAGGAAATTGAACAGATGAATA ATTTGTTCCAGCAAAAACAGAGGGAGCTCGTCCTGGCTGTGTCAAAAGTAGAAGAACTGACCAGGCAGCT AGAGATGCTCAAGAACGGCAGGATCGACAGCCACCATGACAATCAGTCTGCAGTGGCTGAGCTTGATCGC CTCTATAAGGAGCTGCAGCTAAGAAACAAATTGAATCAAGAGCAGAATGCCAAGCTACAACAACAGAGGG AGTGTTTGAATAAGCGTAATTCAGAAGTGGCAGTCATGGATAAGCGTGTTAATGAGCTGAGGGACCGGCT GTGGAAGAAGAAGGCAGCTCTACAGCAAAAAGAAAATCTACCAGTTTCATCTGATGGAAATCTTCCCCAG CAAGCCGCGTCAGCCCCAAGCCGTGTGGCTGCAGTAGGTCCCTATATCCAGTCATCTACTATGCCTCGGA TGCCCTCAAGGCCTGAATTGCTGGTGAAGCCAGCCCTGCCGGATGGTTCCTTGGTCATTCAGGCTTCAGA GGGGCCGATGAAAATACAGACACTGCCCAACATGAGATCTGGGGCTGCTTCACAAACTAAAGGCTCTAAA ATCCATCCAGTTGGCCCTGATTGGAGTCCTTCAAATGCAGATCTTTTCCCAAGCCAAGGCTCTGCTTCTG TACCTCAAAGCACTGGGAATGCTCTGGATCAAGTTGATGATGGAGAGGTTCCGCTGAGGGAGAAAGAGAA GAAAGTGCGTCCGTTCTCAATGTTTGATGCAGTAGACCAGTCCAATGCCCCACCTTCCTTTGGTACTCTG AGGAAGAACCAGAGCAGTGAAGATATCTTGCGGGATGCTCAGGTTGCAAATAAAAATGTGGCTAAAGTAC CACCTCCTGTTCCTACAAAACCAAAACAGATTAATTTGCCTTATTTTGGACAAACTAATCAGCCACCTTC AGACATTAAGCCAGACGGAAGTTCTCAGCAGTTGTCAACAGTTGTTCCGTCCATGGGAACTAAACCAAAA CCAGCAGGGCAGCAGCCGAGAGTGCTGCTATCTCCCAGCATACCTTCGGTTGGCCAAGACCAGACCCTTT CTCCAGGTTCTAAGCAAGAAAGTCCACCTGCTGCTGCCGTCCGGCCCTTTACTCCCCAGCCTTCCAAAGA CACCTTACTTCCACCCTTCAGAAAACCCCAGACCGTGGCAGCAAGTTCAATATATTCCATGTATACGCAA CAGCAGGCGCCAGGAAAAAACTTCCAGCAGGCTGTGCAGAGCGCGTTGACCAAGACTCATACCAGAGGGC CACACTTTTCAAGTGTATATGGTAAGCCTGTAATTGCTGCTGCCCAGAATCAACAGCAGCACCCAGAGAA CATTTATTCCAATAGCCAGGGCAAGCCTGGCAGTCCAGAACCTGAAACAGAGCCTGTTTCTTCAGTTCAG

	GAGAACCATGAAAACGAAAGAATTCCTCGGCCACTCAGCCCAACTAAATTACTGCCTTTCTTATCTAATC CTTACCGAAACCAGAGTGATGCTGACCTAGAAGCCTTACGAAAGAAACTGTCTAACGCACCAAGGCCTCT AAAGAAACGTAGTTCTATTACAGAGCCAGAGGGTCCTAATGGGCCAAATATTCAGAAGCTTTTATATCAG AGGACCACCATAGCGGCCATGGAGACCATCTCTGTCCCATCATACCCATCCAAGTCAGCTTCTGTGACTG CCAGCTCAGAAAGCCCAGTAGAAATCCAGAATCCACATGTGCTTGATGTGTACCTGGAGGAGTACCCTCC ATACCCACCCCCACCATACCCATCTGGGGAGCCTGAAGGGCCCGGAGAAGACTCGGTGAGCATGCGCCCG CCTGAAATCACCGGGCAGGTCTCTCTGCCTCCTGGTAAAAGGACAAACTTGCGTAAAACTGGCTCAGAGC GTATCGCTCATGGAATGAGGGTGAAATTCAACCCCCTTGCTTTACTGCTAGATTCGTCTTTGGAGGGAGA ATTTGACCTTGTACAGAGAATTATTTATGAGGTTGATGACCCAAGCCTGCCCAATGATGAAGGCATCACG GCTCTTCACAATGCTGTGTGTGCAGGCCACACAGAAATCGTTAAGTTCCTGGTACAGTTTGGTGTAAATG TAAATGCTGCTGATAGTGATGGATGGACTCCATTACATTGTGCTGCCTCATGTAACAACGTCCAAGTGTG TAAGTTTTTGGTGGAGTCAGGAGCCGCTGTGTTTGCCATGACCTACAGTGACATGCAGACTGCTGCAGAT AAGTGCGAGGAAATGGAGGAAGGCTACACTCAGTGCTCCCAATTTCTTTATGGAGTTCAGGAGAAGATGG GCATAATGAATAAAGGAGTCATTTATGCGCTTTGGGATTATGAACCTCAGAATGATGATGAGCTGCCCAT GAAAGAAGGAGACTGCATGACAATCATCCACAGGGAAGACGAAGATGAAATCGAATGGTGGTGGGCGCGC CTTAATGATAAGGAGGGATATGTTCCACGTAACTTGCTGGGACTGTACCCAAGAATTAAACCAAGACAAA GGAGCTTGGCCTGAAACTTCCACACAGAATTTTAGTCAATGAAGAATTAATCTCTGTTAAGAAGAAGTAA TACGATTATTTTTGGCAAAAATTTCACAAGACTTATTTTAATGACAATGTAGCTTGAAAGCGATGAAGAA TGTCTCTAGAAGAGAATGAAGGATTGAAGAATTCACCATTAGAGGACATTTAGCGTGATGAAATAAAGCA TCTACGTCAGCAGGCCATACTGTGTTGGGGCAAAGGTGTCCCGTGTAGCACTCAGATAAGTATACAGCGA CAATCCTGTTTTCTACAAGAATCCTGTCTAGTAAATAGGATCATTTATTGGGCAGTTGGGAAATCAGCTC TCTGTCCTGTTGAGTGTTTTCAGCAGCTGCTCCTAAACCAGTCCTCCTGCCAGAAAGGACCAGTGCCGTC ACATCGCTGTCTCTGATTGTCCCCGGCACCAGCAGGCCTTGGGGCTCACTGAAGGCTCGAAGGCACTGCA CACCTTGTATATTGTCAGTGAAGAACGTTAGTTGGTTGTCAGTGAACAATAACTTTATTATATGAGTTTT TGTAGCATCTTAAGAATTATACATATGTTTGAAATATTGAAACTAAGCTACAGTACCAGTAATTAGATGT AGAATCTTGTTTGTAGGCTGAATTTTAATCTGTATTTATTGTCTTTTGTATCTCAGAAATTAGAAACTTG CTACAGACTTACCCGTAATATTTGTCAAGATCATAGCTGACTTTAAAAACAGTTGTAATAAACTTTTTGA TGCT
	ORF Start: ATG at 238 | | ORF Stop: TGA at 3442
	SEQ ID NO: 64 | 1068 aa | MW at 119273.7kD
NOV20b, CG59323-03 Protein Sequence	MRFGSKMMPMFLTVYLSNNEQHFTEVPVTPETICRDVVDLCKEPGESDCHLAEVWCGSERPVADNERMFD VLQRFGSQRNEVRFFLRHERPPGRDIVSGPRSQDPSLKRNGVKVPGEYRRKENGVNSPRMDLTLAELQEM ASRQQQQIEAQQQLLATKEQRLKFLKQQDQRQQQQVAEQEKLKRLKEIAENQEAKLKKVRALKGHVEQKR LSNGKLVEEIEQMNNLFQQKQRELVLAVSKVEELTRQLEMLKNGRIDSHHDNQSAVAELDRLYKELQLRN KLNQEQNAKLQQQRECLNKRNSEVAVMDKRVNELRDRLWKKKAALQQKENLPVSSDGNLPQQAASAPSRV AAVGPYIQSSTMPRMPSRPELLVKPALPDGSLVIQASEGPMKIQTLPNMRSGAASQTKGSKIHPVGPDWS PSNADLFPSQGSASVPQSTGNALDQVDDGEVPLREKEKKVRPFSMFDAVDQSNAPPSFGTLRKNQSSEDI LRDAQVANKNVAKVPPPVPTKPKQINLPYFGQTNQPPSDIKPDGSSQQLSTVVPSMGTKPKPAGQQPRVL LSPSIPSVGQDQTLSPGSKQESPPAAAVRPFTPQPSKDTLLPPFRKPQTVAASSIYSMYTQQQAPGKNFQ QAVQSALTKTHTRGPHFSSVYGKPVIAAAQNQQQHPENIYSNSQGKPGSPEPETEPVSSVQENHENERIP RPLSPTKLLPFLSNPYRNQSDADLEALRKKLSNAPRPLKKRSSITEPEGPNGPNIQKLLYQRTTIAAMET ISVPSYPSKSASVTASSESPVEIQNPHVLDVYLEEYPPYPPPPYPSGEPEGPGEDSVSMRPPEITGQVSL PPGKRTNLRKTGSERIAHGMRVKFNPLALLLDSSLEGEFDLVQRIIYEVDDPSLPNDEGITALHNAVCAG HTEIVKFLVQFGVNVNAADSDGWTPLHCAASCNNVQVCKFLVESGAAVFAMTYSDMQTAADKCEEMEEGY TQCSQFLYGVQEKMGIMNKGVIYALWDYEPQNDDELPMKEGDCMTIIHREDEDEIEWWWARLNDKEGYVP RNLLGLYPRIKPRQRSLA
	SEQ ID NO: 65 | 4336 bp |
NOV20c, CG59323-02 DNA Sequence	GTCACGAGCGTCGAAGAGACAAAGCCGCGTCAGGGGGCCCGGCCGGGGCGGGGGAGCCCGGGGCTTGTTG GTGCCCCAGCCCGCGCGGAGGGCCCTTCGGACCCGCGCGCCGCCGCTGCCGCCGCCGCCGCCTCGCAACA GGTCCGGGCGGCCTCGCTCTCCGCTCCCCTCCCCCGCATCCGCGACCCTCCGGGGCACCTCAGCTCGGCC GGGGCCGCAGTCTGGCCACCCGCTTCCATGCGGTTCGGGTCCAAGATGATGCCGATGTTTCTTACCGTGT ATCTCAGTAACAATGAGCAGCACTTCACAGAAGTTCCAGTTACTCCAGAAACAATATGCAGAGACGTGGT GGATCTGTGCAAAGAACCCGGCGAGAGTGATTGCCATTTGGCTGAAGTGTGGTGTGGCTCTGTAGAGATA GAGTTTCATCATGTTGGCCAGGATGGTCTCGATCTCCTGACCTTGTGATCCGCCTGCCTCGGCCTCCCAA AGTGCTGGATTACAGGTGTGAGCCACCACGATCAGCCTCTAGTGTTTAAAAAAGAACGTCCAGTTGCGGA TAATGAGCGAATGTTTGATGTTCTTCAACGATTTGGAAGTCAGAGGAACGAAGTTCGCTTCTTCCTTCGT CATGAACGCCCCCCTGGCAGGGACATTGTGAGTGGACCAAGATCTCAGGATCCAAGTTTAAAAAGAAATG GTGTAAAAGTTCCTGGTGAATATCGAAGAAAGGAGAACGGTGTTAATAGTCCTAGGATGGATCTGACTCT TGCTGAACTTCAGGAAATGGCATCTCGCCAGCAGCAACAGATTGAAGCCCAGCAACAATTGCTGGCAACT AAGGAACAGCGCTTAAAGTTTTTGAAACAACAAGATCAGCGACAACAGCAACAAGTTGCTGAGCAGGAGA AACTTAAAAGGCTAAAAGAAATAGCTGAGAATCAGGAAGCTAAGCTAAAAAAAGTGAGAGCACTTAAAGG CCACGTGGAACAGAAGAGACTAAGCAATGGGAAACTTGTGGAGGAAATTGAACAGATGAATAATTTGTTC CAGCAAAAACAGAGGGAGCTCGTCCTGGCTGTGTCAAAAGTAGAAGAACTGACCAGGCAGCTAGAGATGC TCAAGAACGGCAGGATCGACAGCCACCATGACAATCAGTCTGCAGTGGCTGAGCTTGATCGCCTCTATAA

	GGAGCTGCAGCTAAGAAACAAATTGAATCAAGAGCAGAATGCCAAGCTACAACAACAGAGGGAGTGTTTG AATAAGCGTAATTCAGAAGTGGCAGTCATGGATAAGCGTGTTAATGAGCTGAGGGACCGGCTGTGGAAGA AGAAGGCAGCTCTACAGCAAAAAGAAAATCTACCAGTTTCATCTGATGGAAATCTTCCCCAGCAAGCCGC GTCAGCCCCAAGCCGTGTGGCTGCAGTAGGTCCCTATATCCAGTCATCTACTATGCCTCGGATGCCCTCA AGGCCTGAATTGCTGGTGAAGCCAGCCCTGCCGGATGGTTCCTTGGTCATTCAGGCTTCAGAGGGGCCGA TGAAAATACAGACACTGCCCAACATGAGATCTGGGGCTGCTTCACAAACTAAAGGCTCTAAAATCCATCC AGTTGGCCCTGATTGGAGTCCTTCAAATGCAGATCTTTTCCCAAGCCAAGGCTCTGCTTCTGTACCTCAA AGCACTGGGAATGCTCTGGATCAAGTTGATGATGGAGAGGTTCCGCTGAGGGAGAAAGAGAAGAAAGTGC GTCCGTTCTCAATGTTTGATGCAGTAGACCAGTCCAATGCCCCACCTTCCTTTGGTACTCTGAGGAAGAA CCAGAGCAGTGAAGATATCTTGCGGGATGCTCAGGTTGCAAATAAAAATGTGGCTAAAGTACCACCTCCT GTTCCTACAAAACCAAAACAGATTAATTTGCCTTATTTTGGACAAACTAATCAGCCACCTTCAGACATTA AGCCAGACGGAAGTTCTCAGCAGTTGTCAACAGTTGTTCCGTCCATGGGAACTAAACCAAAACCAGCAGG GCAGCAGCCGAGAGTGCTGCTATCTCCCAGCATACCTTCGGTTGGCCAAGACCAGACCCTTTCTCCAGGT TCTAAGCAAGAAAGTCCACCTGCTGCTGCCGTCCGGCCCTTTACTCCCCAGCCTTCCAAAGACACCTTAC TTCCACCCTTCAGAAAACCCCAGACCGTGGCAGCAAGTTCAATATATTCCATGTATACGCAACAGCAGGC GCCAGGAAAAAACTTCCAGCAGGCTGTGCAGAGCGCGTTGACCAAGACTCATACCAGAGGGCCACACTTT TCAAGTGTATATGGTAAGCCTGTAATTGCTGCTGCCCAGAATCAACAGCAGCACCCAGAGAACATTTATT CCAATAGCCAGGGCAAGCCTGGCAGTCCAGAACCTGAAACAGAGCCTGTTTCTTCAGTTCAGGAGAACCA TGAAAACGAAAGAATTCCTCGGCCACTCAGCCCAACTAAATTACTGCCTTTCTTATCTAATCCTTACCGA AACCAGAGTGATGCTGACCTAGAAGCCTTACGAAAGAAACTGTCTAACGCACCAAGGCCTCTAAAGAAAC GTAGTTCTATTACAGAGCCAGAGGGTCCTAATGGGCCAAATATTCAGAAGCTTTTATATCAGAGGACCAC CATAGCGGCCATGGAGACCATCTCTGTCCCATCATACCCATCCAAGTCAGCTTCTGTGACTGCCAGCTCA GAAAGCCCAGTAGAAATCCAGAATCCACATGTGCTTGATGTGTACCTGGAGGAGTACCCTCCATACCCAC CCCCACCATACCCATCTGGGGAGCCTGAAGGGCCCGGAGAAGACTCGGTGAGCATGCGCCCGCCTGAAAT CACCGGGCAGGTCTCTCTGCCTCCTGGTAAAAGGACAAACTTGCGTAAAACTGGCTCAGAGCGTATCGCT CATGGAATGAGGGTGAAATTCAACCCCCTTGCTTTACTGCTAGATTCGTCTTTGGAGGGAGAATTTGACC TTGTACAGAGAATTATTTATGAGGTTGATGACCCAAGCCTGCCCAATGATGAAGGCATCACGGCTCTTCA CAATGCTGTGTGTGCAGGCCACACAGAAATCGTTAAGTTCCTGGTACAGTTTGGTGTAAATGTAAATGCT GCTGATAGTGATGGATGGACTCCATTACATTGTGCTGCCTCATGTAACAACGTCCAAGTGTGTAAGTTTT TGGTGGAGTCAGGAGCCGCTGTGTTTGCCATGACCTACAGTGACATGCAGACTGCTGCAGATAAGTGCGA GGAAATGGAGGAAGGCTACACTCAGTGCTCCCAATTTCTTTATGGAGTTCAGGAGAAGATGGGCATAATG AATAAAGGAGTCATTTATGCGCTTTGGGATTATGAACCTCAGAATGATGATGAGCTGCCCATGAAAGAAG GAGACTGCATGACAATCATCCACAGGGAAGACGAAGATGAAATCGAATGGTGGTGGGCGCGCCTTAATGA TAAGGAGGGATATGTTCCACGTAACTTGCTGGGACTGTACCCAAGAATTAAACCAAGACAAAGGAGCTTG GCCTGAAACTTCCACACAGAATTTTAGTCAATGAAGAATTAATCTCTGTTAAGAAGAAGTAATACGATTA TTTTTGGCAAAAATTTCACAAGACTTATTTTAATGACAATGTAGCTTGAAAGCGATGAAGAATGTCTCTA GAAGAGAATGAAGGATTGAAGAATTCACCATTAGAGGACATTTAGCGTGATGAAATAAAGCATCTACGTC AGCAGGCCATACTGTGTTGGGGCAAAGGTGTCCCGTGTAGCACTCAGATAAGTATACAGCGACAATCCTG TTTTCTACAAGAATCCTGTCTAGTAAATAGGATCATTTATTGGGCAGTTGGGAAATCAGCTCTCTGTCCT GTTGAGTGTTTTCAGCAGCTGCTCCTAAACCAGTCCTCCTGCCAGAAAGGACCAGTGCCGTCACATCGCT GTCTCTGATTGTCCCCGGCACCAGCAGGCCTTGGGGCTCACTGAAGGCTCGAAGGCACTGCACACCTTGT ATATTGTCAGTGAAGAACGTTAGTTGGTTGTCAGTGAACAATAACTTTATTATATGAGTTTTTGTAGCAT CTTAAGAATTATACATATGTTTGAAATATTGAAACTAAGCTACAGTACCAGTAATTAGATGTAGAATCTT GTTTGTAGGCTGAATTTTAATCTGTATTTATTGTCTTTTGTATCTCAGAAATTAGAAACTTGCTACAGAC TTACCCGTAATATTTGTCAAGATCATAGCTGACTTTAAAAACAGTTGTAATAAACTTTTTGATGCT
	ORF Start: ATG at 571		ORF Stop: TGA at 3574
	SEQ ID NO: 66	1001 aa	MW at 111615.1kD
NOV20c, CG59323-02 Protein Sequence	MFDVLQRFGSQRNEVRFFLRHERPPGRDIVSGPRSQDPSLKRNGVKVPGEYRRKENGVNSPRMDLTLAEL QEMASRQQQQIEAQQQLLATKEQRLKFLKQQDQRQQQQVAEQEKLKRLKEIAENQEAKLKKVRALKGHVE QKRLSNGKLVEEIEQMNNLFQQKQRELVLAVSKVEELTRQLEMLKNGRIDSHHDNQSAVAELDRLYKELQ LRNKLNQEQNAKLQQQRECLNKRNSEVAVMDKRVNELRDRLWKKKAALQQKENLPVSSDGNLPQQAASAP SRVAAVGPYIQSSTMPRMPSRPELLVKPALPDGSLVIQASEGPMKIQTLPNMRSGAASQTKGSKIHPVGP DWSPSNADLFPSQGSASVPQSTGNALDQVDDGEVPLREKEKKVRPFSMFDAVDQSNAPPSFGTLRKNQSS EDILRDAQVANKNVAKVPPPVPTKPKQINLPYFGQTNQPPSDIKPDGSSQQLSTVVPSMGTKPKPAGQQP RVLLSPSIPSVGQDQTLSPGSKQESPPAAAVRPFTPQPSKDTLLPPFRKPQTVAASSIYSMYTQQQAPGK NFQQAVQSALTKTHTRGPHFSSVYGKPVIAAAQNQQQHPENIYSNSQGKPGSPEPETEPVSSVQENHENE RIPRPLSPTKLLPFLSNPYRNQSDADLEALRKKLSNAPRPLKKRSSITEPEGPNGPNIQKLLYQRTTIAA METISVPSYPSKSASVTASSESPVEIQNPHVLDVYLEEYPPYPPPPYPSGEPEGPGEDSVSMRPPEITGQ VSLPPGKRTNLRKTGSERIAHGMRVKFNPLALLLDSSLEGEFDLVQRIIYEVDDPSLPNDEGITALHNAV CAGHTEIVKFLVQFGVNVNAADSDGWTPLHCAASCNNVQVCKFLVESGAAVFAMTYSDMQTAADKCEEME EGYTQCSQFLYGVQEKMGIMNKGVIYALWDYEPQNDDELPMKEGDCMTIIHREDEDEIEWWWARLNDKEG YVPRNLLGLYPRIKPRQRSLA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b and NOV20c.		
Protein Sequence	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**
NOV20b	1..1005 130..1068	671/1020 (65%) 710/1020 (68%)
NOV20c	1..1005 63..1001	671/1020 (65%) 710/1020 (68%)

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a	
PSort analysis:	0.7600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV20a protein against the GENESEQ database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. GENESEQ Results for NOV20a				
GENESEQ Identifier	**Protein/Organism/Length [Patent #, Date]**	**NOV20a Residues/ Match Residues**	**Identities/ Similarities for the Matched Region**	**Expect Value**
AAU78283	Human apoptosis stimulating protein 2 (APS-2) - Homo sapiens, 1467 aa. [WO200212325-A2, 14-FEB-2002]	1..1005 209..1213	1005/1005 (100%) 1005/1005 (100%)	0.0
AAW93955	Human 53BP2 protein - Homo sapiens, 1005 aa. [WO9915657-A2, 01-APR-1999]	1..1005 1..1005	1005/1005 (100%) 1005/1005 (100%)	0.0
AAM78805	Human protein SEQ ID NO 1467 - Homo sapiens, 1096 aa. [WO200157190-A2, 09-AUG-2001]	1..1005 124..1096	479/1030 (46%) 613/1030 (59%)	0.0
AAU78282	Human apoptosis stimulating protein 1 (APS-1) - Homo sapiens, 1609 aa. [WO200212325-A2, 14-FEB-2002]	1..1005 176..1142	478/1030 (46%) 611/1030 (58%)	0.0

AAM39292	Human polypeptide SEQ ID NO 2437 - Homo sapiens, 1090 aa. [WO200153312-A1, 26-JUL-2001]	1..1005 124..1090	478/1030 (46%) 611/1030 (58%)	0.0

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

Table 20E. Public BLASTP Results for NOV20a

Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96KQ3	ASPP2 protein - Homo sapiens (Human), 1128 aa.	1..1005 124..1128	1005/1005 (100%) 1005/1005 (100%)	0.0
Q13625	Tumor suppressor p53-binding protein 2 (p53-binding protein 2) (53BP2) (Bcl2-binding protein) (Bbp) - Homo sapiens (Human), 1005 aa.	1..1005 1..1005	1005/1005 (100%) 1005/1005 (100%)	0.0
AAH30894	Similar to tumor protein p53 binding protein, 2 - Mus musculus (Mouse), 762 aa (fragment).	258..1005 15..762	664/749 (88%) 688/749 (91%)	0.0
I38607	p53-binding protein 2 - human, 529 aa (fragment).	477..1005 1..529	529/529 (100%) 529/529 (100%)	0.0
Q96KQ4	ASPP1 protein (KIAA0771 protein) - Homo sapiens (Human), 1090 aa.	1..1005 124..1090	478/1030 (46%) 611/1030 (58%)	0.0

PFam analysis predicts that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a

Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ank	835..867	16/33 (48%) 28/33 (85%)	1e-09
ank	868..900	16/33 (48%) 28/33 (85%)	5.1e-09
SH3	937..994	23/61 (38%) 47/61 (77%)	1.2e-12

Example B: Sequencing Methodology and Identification of NOVX Clones

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly

when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3. **PathCalling™ Technology:** The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E. coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly

represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

4. **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5. **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from

exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

Example C. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoinflammatory diseases), Panel CNSD.01 (containing samples from

normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β-actin and GAPDH). Normalized RNA (5 ul) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μg of total RNA were performed in a volume of 20 μl and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μg of total RNA in a final volume of 100 μl. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version 1 for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60°C, primer optimal Tm = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe Tm must be 10°C greater than primer Tm, amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney,

adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

General_screening_panel_v1.4, v1.5 and v1.6

The plates for Panels 1.4, 1.5, and 1.6 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4, 1.5, and 1.6 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, 1.5, and 1.6 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panels 1.4, 1.5, and 1.6 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D, 2.2, 2.3 and 2.4

The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardais or Clinomics). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/ CHTN/Ardais/Clinomics). Unmatched RNA samples from tissues without malignancy (normal tissues) were also obtained from Ardais or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

HASS Panel v 1.0

The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as controls for these treatments, 3 samples of human primary cells, 9 samples of malignant brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have been previously published in the scientific literature. The primary human cells were obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions

recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples . RNA was prepared from these samples using the standard procedures. The genomic and chemistry control wells have been described previously.

ARDAIS Panel v 1.0

The plates for ARDAIS panel v 1.0 generally include 2 control wells and 22 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardais Corporation. The tissues are derived from human lung malignancies (lung adenocarcinoma or lung squamous cell carcinoma) and in cases where indicated many malignant samples have "matched margins" obtained from noncancerous lung tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue) in the results below. The tumor tissue and the "matched margins" are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardais). Unmatched malignant and non-malignant RNA samples from lungs were also obtained from Ardais. Additional information from Ardais provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

Panel 3D, 3.1 and 3.2

The plates of Panel 3D, 3.1, and 3.2 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D, 3.1, 3.2, 1, 1.1., 1.2, 1.3D, 1.4, 1.5, and 1.6 are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final

concentration of approximately $2x10^6$cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol ($5.5x10^{-5}$M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and plated at 10^6cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5μg/ml anti-CD28 (Pharmingen) and 3ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with

plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5μg/ml or anti-CD40 (Pharmingen) at approximately 10μg/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10μg/ml anti-CD28 (Pharmingen) and 2μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5-10^6cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1μg/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1μg/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1μg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at $5x10^5$cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to $5x10^5$cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1µg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5x10^{-5}$M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300µl of RNase-free water and 35µl buffer (Promega) 5µl DTT, 7µl RNAsin and 8µl DNase were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80°C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohn's disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebvid and two were on phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity

Syn = Synovial

Normal = No apparent disease

Rep22 /Rep20 = individual patients

RA = Rheumatoid arthritis

Backus = From Backus Hospital

OA = Osteoarthritis

(SS) (BA) (MF) = Individual patients

Adj = Adjacent tissue

Match control = adjacent tissues

-M = Male

-F = Female

COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2: Diabetic Hispanic, overweight, not on insulin

Patient 7-9: Nondiabetic Caucasian and obese (BMI>30)

Patient 10: Diabetic Hispanic, overweight, on insulin

Patient 11: Nondiabetic African American and overweight

Patient 12: Diabetic Hispanic on insulin

Adiocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U, which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells

Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose

Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated

Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

PL = Placenta

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like

pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; patient not demented but showing sever AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

A. NOV1 CG116579-02: Intracellular protein-like protein.

Expression of gene CG116579-02 was assessed using the primer-probe set Ag6031, described in Table AA. Results of the RTQ-PCR runs are shown in Tables AB, AC, AD and AE.

Table AA. Probe Name Ag6031

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gaccgcaggtgtcgtaaaag-3'	20	162	67
Probe	TET-5'-agtagcccattttggataccgtcctcgc-3'-TAMRA	28	213	68
Reverse	5'-aacgaccagactagcaacaggt-3'	22	258	69

Table AB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag6031, Run 228738344	issue Name	Rel. Exp.(%) Ag6031, Run 228738344
Adipose	1.7	Renal ca. TK-10	15.1
Melanoma* Hs688(A).T	6.4	Bladder	4.7
Melanoma* Hs688(B).T	6.9	Gastric ca. (liver met.) NCI-N87	12.6
Melanoma* M14	14.1	Gastric ca. KATO III	**100.0**
Melanoma* LOXIMVI	21.5	Colon ca. SW-948	10.4
Melanoma* SK-MEL-5	4.9	Colon ca. SW480	37.9
Squamous cell carcinoma SCC-4	6.4	Colon ca.* (SW480 met) SW620	26.2
Testis Pool	4.1	Colon ca. HT29	24.1
Prostate ca.* (bone met) PC-3	10.7	Colon ca. HCT-116	54.0
Prostate Pool	2.0	Colon ca. CaCo-2	21.0
Placenta	4.8	Colon cancer tissue	5.2
Uterus Pool	0.6	Colon ca. SW1116	11.3
Ovarian ca. OVCAR-3	24.0	Colon ca. Colo-205	16.6
Ovarian ca. SK-OV-3	26.6	Colon ca. SW-48	7.2
Ovarian ca. OVCAR-4	17.2	Colon Pool	2.8
Ovarian ca. OVCAR-5	10.7	Small Intestine Pool	1.3
Ovarian ca. IGROV-1	11.2	Stomach Pool	0.9
Ovarian ca. OVCAR-8	7.7	Bone Marrow Pool	1.3
Ovary	1.7	Fetal Heart	1.1
Breast ca. MCF-7	41.8	Heart Pool	0.4
Breast ca. MDA-MB-231	24.7	Lymph Node Pool	3.8
Breast ca. BT 549	27.4	Fetal Skeletal Muscle	1.3
Breast ca. T47D	21.8	Skeletal Muscle Pool	3.2
Breast ca. MDA-N	13.3	Spleen Pool	3.6
Breast Pool	2.5	Thymus Pool	5.2
Trachea	4.8	CNS cancer (glio/astro) U87-MG	9.6
Lung	0.6	CNS cancer (glio/astro) U-118-MG	16.2
Fetal Lung	11.3	CNS cancer (neuro;met) SK-N-AS	9.3
Lung ca. NCI-N417	16.8	CNS cancer (astro) SF-539	15.8
Lung ca. LX-1	15.2	CNS cancer (astro) SNB-75	28.9
Lung ca. NCI-H146	15.2	CNS cancer (glio) SNB-19	11.0
Lung ca. SHP-77	11.2	CNS cancer (glio) SF-295	10.8
Lung ca. A549	18.4	Brain (Amygdala) Pool	10.0
Lung ca. NCI-H526	9.8	Brain (cerebellum)	9.2
Lung ca. NCI-H23	16.3	Brain (fetal)	8.8
Lung ca. NCI-H460	17.2	Brain (Hippocampus) Pool	5.0
Lung ca. HOP-62	9.0	Cerebral Cortex Pool	6.2
Lung ca. NCI-H522	11.3	Brain (Substantia nigra) Pool	6.3

Liver	0.6	Brain (Thalamus) Pool	9.0
Fetal Liver	10.0	Brain (whole)	9.8
Liver ca. HepG2	6.3	Spinal Cord Pool	2.3
Kidney Pool	3.3	Adrenal Gland	4.6
Fetal Kidney	3.4	Pituitary gland Pool	1.6
Renal ca. 786-0	9.9	Salivary Gland	2.0
Renal ca. A498	2.4	Thyroid (female)	2.9
Renal ca. ACHN	2.8	Pancreatic ca. CAPAN2	15.8
Renal ca. UO-31	4.0	Pancreas Pool	5.4

Table AC. Oncology_cell_line_screening_panel_v3.1

Tissue Name	Rel. Exp.(%) Ag6031, Run 226203293	Tissue Name	Rel. Exp.(%) Ag6031, Run 226203293
Daoy Medulloblastoma/Cerebellum	4.9	Ca Ski_Cervical epidermoid carcinoma (metastasis)	34.9
TE671 Medulloblastom/Cerebellum	7.2	ES-2_Ovarian clear cell carcinoma	20.7
D283 Med Medulloblastoma/Cerebellum	10.4	Ramos/6h stim_ Stimulated with PMA/ionomycin 6h	35.4
PFSK-1 Primitive Neuroectodermal/Cerebellum	36.3	Ramos/14h stim_ Stimulated with PMA/ionomycin 14h	20.0
XF-498_CNS	33.4	MEG-01_Chronic myelogenous leukemia (megokaryoblast)	30.1
SNB-78_CNS/glioma	10.7	Raji_Burkitt's lymphoma	4.4
SF-268_CNS/glioblastoma	15.9	Daudi_Burkitt's lymphoma	23.5
T98G_Glioblastoma	17.4	U266_B-cell plasmacytoma/myeloma	7.5
SK-N-SH_Neuroblastoma (metastasis)	39.0	CA46_Burkitt's lymphoma	0.0
SF-295_CNS/glioblastoma	6.6	RL_non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	5.4	JM1_pre-B-cell lymphoma/leukemia	9.1
Cerebellum	3.8	Jurkat_T cell leukemia	25.2
NCI-H292_Mucoepidermoid lung ca.	16.8	TF-1_Erythroleukemia	20.6
DMS-114_Small cell lung cancer	10.7	HUT 78_T-cell lymphoma	57.0
DMS-79_Small cell lung cancer/neuroendocrine	22.7	U937_Histiocytic lymphoma	82.9
NCI-H146_Small cell lung cancer/neuroendocrine	58.6	KU-812_Myelogenous leukemia	9.9
NCI-H526_Small cell lung cancer/neuroendocrine	59.5	769-P_Clear cell renal ca.	10.3
NCI-N417_Small cell lung cancer/neuroendocrine	35.6	Caki-2_Clear cell renal ca.	12.2

NCI-H82_Small cell lung cancer/neuroendocrine	6.0	SW 839_Clear cell renal ca.	7.3
NCI-H157_Squamous cell lung cancer (metastasis)	25.0	G401_Wilms' tumor	8.7
NCI-H1155_Large cell lung cancer/neuroendocrine	71.2	Hs766T_Pancreatic ca. (LN metastasis)	21.6
NCI-H1299_Large cell lung cancer/neuroendocrine	79.6	CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)	9.3
NCI-H727_Lung carcinoid	44.4	SU86.86_Pancreatic carcinoma (liver metastasis)	22.7
NCI-UMC-11_Lung carcinoid	39.5	BxPC-3_Pancreatic adenocarcinoma	11.5
LX-1_Small cell lung cancer	17.0	HPAC_Pancreatic adenocarcinoma	16.0
Colo-205_Colon cancer	38.7	MIA PaCa-2_Pancreatic ca.	4.5
KM12_Colon cancer	40.1	CFPAC-1_Pancreatic ductal adenocarcinoma	55.9
KM20L2_Colon cancer	16.5	PANC-1_Pancreatic epithelioid ductal ca.	15.7
NCI-H716_Colon cancer	36.6	T24_Bladder ca. (transitional cell)	9.9
SW-48_Colon adenocarcinoma	20.9	5637_Bladder ca.	7.8
SW1116_Colon adenocarcinoma	10.9	HT-1197_Bladder ca.	27.4
LS 174T_Colon adenocarcinoma	31.9	UM-UC-3_Bladder ca. (transitional cell)	3.1
SW-948_Colon adenocarcinoma	12.7	A204_Rhabdomyosarcoma	4.9
SW-480_Colon adenocarcinoma	20.6	HT-1080_Fibrosarcoma	7.0
NCI-SNU-5_Gastric ca.	24.1	MG-63_Osteosarcoma (bone)	6.6
KATO III_Stomach	**100.0**	SK-LMS-1_Leiomyosarcoma (vulva)	18.0
NCI-SNU-16_Gastric ca.	19.6	SJRH30_Rhabdomyosarcoma (met to bone marrow)	30.4
NCI-SNU-1_Gastric ca.	9.5	A431_Epidermoid ca.	76.8
RF-1_Gastric adenocarcinoma	6.7	WM266-4_Melanoma	13.5
RF-48_Gastric adenocarcinoma	4.7	DU 145_Prostate	28.9
MKN-45_Gastric ca.	30.1	MDA-MB-468_Breast adenocarcinoma	7.0
NCI-N87_Gastric ca.	1.2	SSC-4_Tongue	11.2
OVCAR-5_Ovarian ca.	3.6	SSC-9_Tongue	73.7
RL95-2_Uterine carcinoma	15.4	SSC-15_Tongue	28.7
HelaS3_Cervical adenocarcinoma	8.2	CAL 27_Squamous cell ca. of tongue	10.4

Table AD. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag6031, Run 225428031	Tissue Name	Rel. Exp.(%) Ag6031, Run 225428031

Secondary Th1 act	7.3	HUVEC IL-1beta	3.8
Secondary Th2 act	3.9	HUVEC IFN gamma	2.1
Secondary Tr1 act	5.1	HUVEC TNF alpha + IFN gamma	1.3
Secondary Th1 rest	1.6	HUVEC TNF alpha + IL4	5.9
Secondary Th2 rest	6.6	HUVEC IL-11	0.7
Secondary Tr1 rest	6.7	Lung Microvascular EC none	5.1
Primary Th1 act	2.8	Lung Microvascular EC TNFalpha + IL-1beta	1.4
Primary Th2 act	6.5	Microvascular Dermal EC none	1.3
Primary Tr1 act	3.5	Microvascular Dermal EC TNFalpha + IL-1beta	0.6
Primary Th1 rest	2.6	Bronchial epithelium TNFalpha + IL1beta	0.6
Primary Th2 rest	7.3	Small airway epithelium none	5.1
Primary Tr1 rest	11.6	Small airway epithelium TNFalpha + IL-1beta	2.1
CD45RA CD4 lymphocyte act	2.9	Coronary artery SMC rest	1.3
CD45RO CD4 lymphocyte act	6.0	Coronary artery SMC TNFalpha + IL-1beta	1.6
CD8 lymphocyte act	6.0	Astrocytes rest	1.6
Secondary CD8 lymphocyte rest	6.5	Astrocytes TNFalpha + IL-1beta	1.1
Secondary CD8 lymphocyte act	5.0	KU-812 (Basophil) rest	1.7
CD4 lymphocyte none	1.6	KU-812 (Basophil) PMA/ionomycin	3.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	11.0	CCD1106 (Keratinocytes) none	18.3
LAK cells rest	2.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.9
LAK cells IL-2	7.3	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	6.4	NCI-H292 none	1.3
LAK cells IL-2+IFN gamma	4.4	NCI-H292 IL-4	1.3
LAK cells IL-2+ IL-18	10.2	NCI-H292 IL-9	4.5
LAK cells PMA/ionomycin	3.1	NCI-H292 IL-13	2.1
NK Cells IL-2 rest	9.1	NCI-H292 IFN gamma	2.8
Two Way MLR 3 day	1.7	HPAEC none	0.7
Two Way MLR 5 day	2.6	HPAEC TNF alpha + IL-1 beta	0.9
Two Way MLR 7 day	2.5	Lung fibroblast none	0.8
PBMC rest	1.2	Lung fibroblast TNF alpha + IL-1 beta	1.8
PBMC PWM	4.4	Lung fibroblast IL-4	1.0
PBMC PHA-L	5.9	Lung fibroblast IL-9	4.3
Ramos (B cell) none	9.9	Lung fibroblast IL-13	2.1
Ramos (B cell) ionomycin	10.6	Lung fibroblast IFN gamma	2.9
B lymphocytes PWM	5.6	Dermal fibroblast CCD1070 rest	1.0
B lymphocytes CD40L and IL-4	3.3	Dermal fibroblast CCD1070 TNF alpha	7.5

EOL-1 dbcAMP	5.3	Dermal fibroblast CCD1070 IL-1 beta	1.4
EOL-1 dbcAMP PMA/ionomycin	5.1	Dermal fibroblast IFN gamma	3.9
Dendritic cells none	3.1	Dermal fibroblast IL-4	2.4
Dendritic cells LPS	1.5	Dermal Fibroblasts rest	1.7
Dendritic cells anti-CD40	1.4	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.8	Neutrophils rest	1.0
Monocytes LPS	1.6	Colon	1.0
Macrophages rest	2.1	Lung	4.1
Macrophages LPS	0.0	Thymus	20.0
HUVEC none	3.6	Kidney	**100.0**
HUVEC starved	4.1		

Table AE. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag603, Run 253578283	Tissue Name	Rel. Exp.(%) Ag6031, Run 253578283
97457_Patient-02go_adipose	24.8	94709_Donor 2 AM - A_adipose	63.7
97476_Patient-07sk_skeletal muscle	0.0	94710_Donor 2 AM - B_adipose	22.4
97477_Patient-07ut_uterus	16.4	94711_Donor 2 AM - C_adipose	23.5
97478_Patient-07pl_placenta	29.9	94712_Donor 2 AD - A_adipose	**100.0**
99167_Bayer Patient 1	11.4	94713_Donor 2 AD - B_adipose	58.6
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	81.2
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	62.4
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	18.9
97487_Patient-09ut_uterus	57.8	94730_Donor 3 AM - A_adipose	16.8
97488_Patient-09pl_placenta	29.5	94731_Donor 3 AM - B_adipose	49.3
97492_Patient-10ut_uterus	2.0	94732_Donor 3 AM - C_adipose	11.0
97493_Patient-10pl_placenta	36.9	94733_Donor 3 AD - A_adipose	21.2
97495_Patient-11go_adipose	38.7	94734_Donor 3 AD - B_adipose	19.1
97496_Patient-11sk_skeletal muscle	2.1	94735_Donor 3 AD - C_adipose	71.7
97497_Patient-11ut_uterus	38.7	77138_Liver_HepG2untreated	28.9
97498_Patient-11pl_placenta	30.1	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	23.7	81735_Small Intestine	46.0
97501_Patient-12sk_skeletal muscle	8.8	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	12.8	82685_Small intestine_Duodenum	11.2

97503_Patient-12pl_placenta	11.7	90650_Adrenal_Adrenocortical adenoma	7.2
94721_Donor 2 U - A_Mesenchymal Stem Cells	66.0	72410_Kidney_HRCE	82.4
94722_Donor 2 U - B_Mesenchymal Stem Cells	40.6	72411_Kidney_HRE	7.1
94723_Donor 2 U - C_Mesenchymal Stem Cells	47.6	73139_Uterus_Uterine smooth muscle cells	13.7

General_screening_panel_v1.5 Summary: Ag6031 Highest expression of this gene is detected in gastric cancer KATO III cell line (CT=28.9). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at low levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, and fetal liver. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate to low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Oncology_cell_line_screening_panel_v3.1 Summary: Ag6031 Highest levels of expression of this gene is detected in gastric cancer KATO III cell line (CT=31). This gene shows a widespread expression in this panel. Moderate to low levels of expression of this gene is seen in number of cell lines derived from tongue, breast, prostate, melanoma, epidermoid, bone, bone-marrow, vulva, bladder, pancreatic, renal, gastric, colon, lung, and brain cancers. Moderate to low levels of expression of this gene is also seen in Wilm's tumor, histiocytic lymphoma, T cell and B cell lymphoma/leukemia, and Burkitt's lymphoma. Therefore, expression of this gene may be used as diagnostic marker to detect

the presence of all these cancers and also, therapeutic modulation of this gene or its protein product may be useful in the treatment of these cancers.

Panel 4.1D Summary: Ag6031 Highest expression of this gene is detected in kidney (CT=29.1). This gene is expressed at low to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Panel 5 Islet Summary: Ag6031 Highest expression of this gene is detected in differentiated adipose tissue (CT=32.5). Moderate to low levels of expression of this gene is also seen in placenta, uterus, small intestine and kidney samples. Please see panel 1.5 for further discussion on the utility of this gene.

B. NOV3 CG137623-01: 2310038H17RIK protein-like protein (TmSP).

Expression of gene CG137623-01 was assessed using the primer-probe set Ag4919, described in Table BA. Results of the RTQ-PCR runs are shown in Tables BB, BC and BD.

Table BA. Probe Name Ag4919

Primers		Length	Start Position	SEQ ID No
Forward	5'-gactggagacacctgggagt-3'	20	322	70
Probe	TET-5'-gatgcggccatcgttctttccac-3'-TAMRA	23	358	71
Reverse	5'-cagctcccatctccaggtat-3'	20	381	72

Table BB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag4919, Run 224997650	issue Name	Rel. Exp.(%) Ag4919, Run 224997650
AD 1 Hippo	8.4	Control (Path) 3 Temporal Ctx	9.6
AD 2 Hippo	18.6	Control (Path) 4 Temporal Ctx	19.6
AD 3 Hippo	12.2	AD 1 Occipital Ctx	14.4
AD 4 Hippo	4.1	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	73.2	AD 3 Occipital Ctx	15.0
AD 6 Hippo	44.1	AD 4 Occipital Ctx	12.3
Control 2 Hippo	28.9	AD 5 Occipital Ctx	16.3
Control 4 Hippo	15.4	AD 6 Occipital Ctx	33.4
Control (Path) 3 Hippo	6.3	Control 1 Occipital Ctx	4.3
AD 1 Temporal Ctx	14.4	Control 2 Occipital Ctx	62.4
AD 2 Temporal Ctx	15.6	Control 3 Occipital Ctx	11.2
AD 3 Temporal Ctx	9.5	Control 4 Occipital Ctx	7.1
AD 4 Temporal Ctx	14.8	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	82.4	Control (Path) 2 Occipital Ctx	5.8
AD 5 SupTemporal Ctx	39.5	Control (Path) 3 Occipital Ctx	5.2
AD 6 Inf Temporal Ctx	48.0	Control (Path) 4 Occipital Ctx	15.5
AD 6 Sup Temporal Ctx	42.9	Control 1 Parietal Ctx	6.1
Control 1 Temporal Ctx	4.8	Control 2 Parietal Ctx	25.7
Control 2 Temporal Ctx	42.9	Control 3 Parietal Ctx	9.0
Control 3 Temporal Ctx	12.4	Control (Path) 1 Parietal Ctx	64.2
Control 4 Temporal Ctx	10.2	Control (Path) 2 Parietal Ctx	30.6
Control (Path) 1 Temporal Ctx	44.8	Control (Path) 3 Parietal Ctx	4.1
Control (Path) 2 Temporal Ctx	24.5	Control (Path) 4 Parietal Ctx	38.7

Table BC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4919, Run 222456686	issue Name	Rel. Exp.(%) Ag4919, Run 222456686
Adipose	5.1	Renal ca. TK-10	70.7
Melanoma* Hs688(A).T	8.2	Bladder	12.5
Melanoma* Hs688(B).T	6.6	Gastric ca. (liver met.) NCI-N87	95.9
Melanoma* M14	37.6	Gastric ca. KATO III	81.8
Melanoma* LOXIMVI	24.0	Colon ca. SW-948	16.6
Melanoma* SK-MEL-5	32.1	Colon ca. SW480	89.5
Squamous cell carcinoma SCC-4	12.2	Colon ca.* (SW480 met) SW620	46.7
Testis Pool	5.4	Colon ca. HT29	34.2

Prostate ca.* (bone met) PC-3	32.3	Colon ca. HCT-116	76.8
Prostate Pool	7.0	Colon ca. CaCo-2	68.3
Placenta	3.7	Colon cancer tissue	13.7
Uterus Pool	4.2	Colon ca. SW1116	12.3
Ovarian ca. OVCAR-3	20.7	Colon ca. Colo-205	12.5
Ovarian ca. SK-OV-3	58.2	Colon ca. SW-48	18.7
Ovarian ca. OVCAR-4	28.9	Colon Pool	11.8
Ovarian ca. OVCAR-5	50.7	Small Intestine Pool	18.4
Ovarian ca. IGROV-1	25.0	Stomach Pool	4.6
Ovarian ca. OVCAR-8	5.3	Bone Marrow Pool	3.5
Ovary	8.2	Fetal Heart	14.6
Breast ca. MCF-7	44.4	Heart Pool	5.6
Breast ca. MDA-MB-231	23.7	Lymph Node Pool	10.7
Breast ca. BT 549	64.2	Fetal Skeletal Muscle	6.7
Breast ca. T47D	77.9	Skeletal Muscle Pool	14.5
Breast ca. MDA-N	41.2	Spleen Pool	10.0
Breast Pool	8.7	Thymus Pool	22.2
Trachea	12.0	CNS cancer (glio/astro) U87-MG	70.2
Lung	4.9	CNS cancer (glio/astro) U-118-MG	**100.0**
Fetal Lung	14.1	CNS cancer (neuro;met) SK-N-AS	59.5
Lung ca. NCI-N417	12.3	CNS cancer (astro) SF-539	25.7
Lung ca. LX-1	51.4	CNS cancer (astro) SNB-75	23.5
Lung ca. NCI-H146	13.4	CNS cancer (glio) SNB-19	14.9
Lung ca. SHP-77	79.6	CNS cancer (glio) SF-295	36.6
Lung ca. A549	41.5	Brain (Amygdala) Pool	12.5
Lung ca. NCI-H526	11.7	Brain (cerebellum)	17.8
Lung ca. NCI-H23	92.0	Brain (fetal)	6.4
Lung ca. NCI-H460	45.7	Brain (Hippocampus) Pool	6.4
Lung ca. HOP-62	14.1	Cerebral Cortex Pool	12.0
Lung ca. NCI-H522	95.9	Brain (Substantia nigra) Pool	8.8
Liver	0.9	Brain (Thalamus) Pool	19.5
Fetal Liver	15.1	Brain (whole)	6.0
Liver ca. HepG2	54.3	Spinal Cord Pool	15.6
Kidney Pool	14.3	Adrenal Gland	5.5
Fetal Kidney	17.9	Pituitary gland Pool	7.9
Renal ca. 786-0	39.0	Salivary Gland	3.4
Renal ca. A498	22.8	Thyroid (female)	9.2
Renal ca. ACHN	31.4	Pancreatic ca. CAPAN2	40.9
Renal ca. UO-31	23.3	Pancreas Pool	23.0

Table BD. Panel 4.1D

Tissue Name	Rel. Exp.() Ag4919, Run 223458648	Tissue Name	Rel. Exp.(%) Ag4919, Run 223458648
Secondary Th1 act	29.1	HUVEC IL-1beta	14.5
Secondary Th2 act	29.1	HUVEC IFN gamma	8.2
Secondary Tr1 act	20.0	HUVEC TNF alpha + IFN gamma	5.4
Secondary Th1 rest	4.6	HUVEC TNF alpha + IL4	13.6
Secondary Th2 rest	12.8	HUVEC IL-11	6.4
Secondary Tr1 rest	5.0	Lung Microvascular EC none	21.2
Primary Th1 act	29.7	Lung Microvascular EC TNFalpha + IL-1beta	13.8
Primary Th2 act	46.7	Microvascular Dermal EC none	12.8
Primary Tr1 act	59.9	Microvascular Dermal EC TNFalpha + IL-1beta	12.1
Primary Th1 rest	11.1	Bronchial epithelium TNFalpha + IL1beta	9.4
Primary Th2 rest	5.4	Small airway epithelium none	2.7
Primary Tr1 rest	17.2	Small airway epithelium TNFalpha + IL-1beta	4.7
CD45RA CD4 lymphocyte act	16.7	Coronary artery SMC rest	8.5
CD45RO CD4 lymphocyte act	33.0	Coronary artery SMC TNFalpha + IL-1beta	2.5
CD8 lymphocyte act	28.7	Astrocytes rest	0.9
Secondary CD8 lymphocyte rest	40.1	Astrocytes TNFalpha + IL-1beta	2.1
Secondary CD8 lymphocyte act	16.0	KU-812 (Basophil) rest	33.7
CD4 lymphocyte none	6.5	KU-812 (Basophil) PMA/ionomycin	52.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	12.8	CCD1106 (Keratinocytes) none	10.9
LAK cells rest	17.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.1
LAK cells IL-2	33.0	Liver cirrhosis	5.9
LAK cells IL-2+IL-12	19.8	NCI-H292 none	28.5
LAK cells IL-2+IFN gamma	17.1	NCI-H292 IL-4	42.3
LAK cells IL-2+ IL-18	14.7	NCI-H292 IL-9	44.1
LAK cells PMA/ionomycin	18.2	NCI-H292 IL-13	11.9
NK Cells IL-2 rest	15.3	NCI-H292 IFN gamma	19.8
Two Way MLR 3 day	19.6	HPAEC none	4.0
Two Way MLR 5 day	33.2	HPAEC TNF alpha + IL-1 beta	6.1
Two Way MLR 7 day	11.8	Lung fibroblast none	12.3
PBMC rest	9.4	Lung fibroblast TNF alpha + IL-1 beta	7.1
PBMC PWM	28.7	Lung fibroblast IL-4	7.5

PBMC PHA-L	46.3	Lung fibroblast IL-9	15.2
Ramos (B cell) none	82.9	Lung fibroblast IL-13	6.3
Ramos (B cell) ionomycin	**100.0**	Lung fibroblast IFN gamma	5.0
B lymphocytes PWM	55.1	Dermal fibroblast CCD1070 rest	5.8
B lymphocytes CD40L and IL-4	40.6	Dermal fibroblast CCD1070 TNF alpha	15.4
EOL-1 dbcAMP	23.8	Dermal fibroblast CCD1070 IL-1 beta	4.5
EOL-1 dbcAMP PMA/ionomycin	9.2	Dermal fibroblast IFN gamma	5.7
Dendritic cells none	29.9	Dermal fibroblast IL-4	14.4
Dendritic cells LPS	8.7	Dermal Fibroblasts rest	5.3
Dendritic cells anti-CD40	12.9	Neutrophils TNFa+LPS	7.7
Monocytes rest	26.4	Neutrophils rest	3.6
Monocytes LPS	12.2	Colon	4.7
Macrophages rest	30.4	Lung	7.4
Macrophages LPS	9.0	Thymus	30.4
HUVEC none	12.7	Kidney	13.8
HUVEC starved	20.2		

CNS_neurodegeneration_v1.0 Summary: Ag4919 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.4 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.4 Summary: Ag4919 Highest expression of this gene is detected in a brain cancer U-118-MG cell line (CT=31.2). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at low levels in pancreas, thyroid, pituitary gland, skeletal muscle, fetal heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may

prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag4919 Highest expression of this gene is detected in ionomycin treated Ramos B cells (CT=32.6). This gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by thymus. This widespread expression pattern suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

C. NOV4 CG137687-01: Cyclin (5730405I09RIK Homolog).

Expression of gene CG137687-01 was assessed using the primer-probe set Ag4926, described in Table CA. Results of the RTQ-PCR runs are shown in Tables CB, CC and CD.

Table CA. Probe Name Ag4926

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gctccttagcagatgacaacaa-3'	22	843	73
Probe	TET-5'-cctgaattttctatttgctcctcttagca-3'-TAMRA	29	865	74
Reverse	5'-agcctctaggttctgtgctctt-3'	22	898	75

Table CB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag4926, Run 224735007	issue Name	Rel. Exp.(%) Ag4926, Run 224735007
AD 1 Hippo	14.5	Control (Path) 3 Temporal Ctx	12.1
AD 2 Hippo	29.7	Control (Path) 4 Temporal Ctx	32.3
AD 3 Hippo	7.0	AD 1 Occipital Ctx	18.4
AD 4 Hippo	10.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	62.0	AD 3 Occipital Ctx	8.0
AD 6 Hippo	56.3	AD 4 Occipital Ctx	31.0
Control 2 Hippo	31.0	AD 5 Occipital Ctx	37.9
Control 4 Hippo	16.8	AD 6 Occipital Ctx	26.2
Control (Path) 3 Hippo	7.1	Control 1 Occipital Ctx	4.7
AD 1 Temporal Ctx	23.2	Control 2 Occipital Ctx	34.6
AD 2 Temporal Ctx	24.7	Control 3 Occipital Ctx	17.3
AD 3 Temporal Ctx	7.5	Control 4 Occipital Ctx	9.0
AD 4 Temporal Ctx	27.5	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	72.7	Control (Path) 2 Occipital Ctx	16.2
AD 5 Sup Temporal Ctx	50.0	Control (Path) 3 Occipital Ctx	5.4
AD 6 Inf Temporal Ctx	70.2	Control (Path) 4 Occipital Ctx	12.7
AD 6 Sup Temporal Ctx	74.7	Control 1 Parietal Ctx	12.0
Control 1 Temporal Ctx	10.4	Control 2 Parietal Ctx	51.4
Control 2 Temporal Ctx	38.2	Control 3 Parietal Ctx	16.4
Control 3 Temporal Ctx	29.1	Control (Path) 1 Parietal Ctx	82.4
Control 3 Temporal Ctx	11.0	Control (Path) 2 Parietal Ctx	28.3
Control (Path) 1 Temporal Ctx	75.8	Control (Path) 3 Parietal Ctx	7.0
Control (Path) 2 Temporal Ctx	49.7	Control (Path) 4 Parietal Ctx	31.9

Table CC. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag4926, Run 228839252	issue Name	Rel. Exp.(%) Ag4926, Run 228839252
Adipose	30.1	Renal ca. TK-10	87.7
Melanoma* Hs688(A).T	32.5	Bladder	28.1
Melanoma* Hs688(B).T	35.6	Gastric ca. (liver met.) NCI-N87	70.7
Melanoma* M14	33.9	Gastric ca. KATO III	52.5
Melanoma* LOXIMVI	36.6	Colon ca. SW-948	11.2
Melanoma* SK-MEL-5	47.6	Colon ca. SW480	72.2
Squamous cell carcinoma SCC-4	28.1	Colon ca.* (SW480 met) SW620	21.3
Testis Pool	18.8	Colon ca. HT29	10.8

Prostate ca.* (bone met) PC-3	33.7	Colon ca. HCT-116	43.2
Prostate Pool	7.7	Colon ca. CaCo-2	**100.0**
Placenta	4.8	Colon cancer tissue	16.8
Uterus Pool	14.5	Colon ca. SW1116	3.7
Ovarian ca. OVCAR-3	36.1	Colon ca. Colo-205	5.7
Ovarian ca. SK-OV-3	55.9	Colon ca. SW-48	5.1
Ovarian ca. OVCAR-4	45.4	Colon Pool	17.8
Ovarian ca. OVCAR-5	74.7	Small Intestine Pool	9.0
Ovarian ca. IGROV-1	13.4	Stomach Pool	8.8
Ovarian ca. OVCAR-8	3.5	Bone Marrow Pool	10.2
Ovary	9.1	Fetal Heart	11.0
Breast ca. MCF-7	31.9	Heart Pool	8.1
Breast ca. MDA-MB-231	35.4	Lymph Node Pool	19.9
Breast ca. BT 549	25.7	Fetal Skeletal Muscle	7.4
Breast ca. T47D	9.2	Skeletal Muscle Pool	17.9
Breast ca. MDA-N	14.7	Spleen Pool	10.8
Breast Pool	11.1	Thymus Pool	9.3
Trachea	16.4	CNS cancer (glio/astro) U87-MG	50.0
Lung	4.9	CNS cancer (glio/astro) U-118-MG	79.6
Fetal Lung	34.2	CNS cancer (neuro;met) SK-N-AS	31.4
Lung ca. NCI-N417	3.5	CNS cancer (astro) SF-539	19.1
Lung ca. LX-1	18.3	CNS cancer (astro) SNB-75	34.6
Lung ca. NCI-H146	2.2	CNS cancer (glio) SNB-19	15.6
Lung ca. SHP-77	41.8	CNS cancer (glio) SF-295	56.6
Lung ca. A549	63.3	Brain (Amygdala) Pool	10.3
Lung ca. NCI-H526	6.7	Brain (cerebellum)	10.6
Lung ca. NCI-H23	44.8	Brain (fetal)	29.7
Lung ca. NCI-H460	57.8	Brain (Hippocampus) Pool	8.7
Lung ca. HOP-62	15.2	Cerebral Cortex Pool	13.6
Lung ca. NCI-H522	52.5	Brain (Substantia nigra) Pool	10.2
Liver	2.5	Brain (Thalamus) Pool	15.5
Fetal Liver	30.6	Brain (whole)	20.6
Liver ca. HepG2	48.0	Spinal Cord Pool	12.2
Kidney Pool	21.6	Adrenal Gland	9.9
Fetal Kidney	17.9	Pituitary gland Pool	2.5
Renal ca. 786-0	38.2	Salivary Gland	8.0
Renal ca. A498	21.6	Thyroid (female)	8.8
Renal ca. ACHN	65.5	Pancreatic ca. CAPAN2	37.9
Renal ca. UO-31	48.0	Pancreas Pool	18.3

Table CD. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag4926, Run 223598854	Tissue Name	Rel. Exp.(%) Ag4926, Run 223598854
Secondary Th1 act	12.9	HUVEC IL-1beta	65.5
Secondary Th2 act	13.8	HUVEC IFN gamma	67.8
Secondary Tr1 act	11.5	HUVEC TNF alpha + IFN gamma	48.6
Secondary Th1 rest	5.4	HUVEC TNF alpha + IL4	40.3
Secondary Th2 rest	7.9	HUVEC IL-11	36.9
Secondary Tr1 rest	6.2	Lung Microvascular EC none	84.7
Primary Th1 act	9.3	Lung Microvascular EC TNFalpha + IL-1beta	48.0
Primary Th2 act	20.6	Microvascular Dermal EC none	52.9
Primary Tr1 act	13.2	Microvascular Dermal EC TNFalpha + IL-1beta	29.5
Primary Th1 rest	5.0	Bronchial epithelium TNFalpha + IL1beta	24.0
Primary Th2 rest	5.6	Small airway epithelium none	13.5
Primary Tr1 rest	10.0	Small airway epithelium TNFalpha + IL-1beta	26.1
CD45RA CD4 lymphocyte act	12.9	Coronary artery SMC rest	26.6
CD45RO CD4 lymphocyte act	14.4	Coronary artery SMC TNFalpha + IL-1beta	21.6
CD8 lymphocyte act	6.1	Astrocytes rest	17.8
Secondary CD8 lymphocyte rest	8.6	Astrocytes TNFalpha + IL-1beta	17.9
Secondary CD8 lymphocyte act	3.3	KU-812 (Basophil) rest	11.5
CD4 lymphocyte none	6.7	KU-812 (Basophil) PMA/ionomycin	15.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	8.1	CCD1106 (Keratinocytes) none	37.6
LAK cells rest	26.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	49.7
LAK cells IL-2	9.1	Liver cirrhosis	6.7
LAK cells IL-2+IL-12	5.7	NCI-H292 none	17.7
LAK cells IL-2+IFN gamma	10.9	NCI-H292 IL-4	57.0
LAK cells IL-2+ IL-18	12.2	NCI-H292 IL-9	57.4
LAK cells PMA/ionomycin	32.8	NCI-H292 IL-13	52.5
NK Cells IL-2 rest	11.0	NCI-H292 IFN gamma	27.2
Two Way MLR 3 day	16.3	HPAEC none	62.0
Two Way MLR 5 day	14.3	HPAEC TNF alpha + IL-1 beta	**100.0**
Two Way MLR 7 day	6.7	Lung fibroblast none	17.4
PBMC rest	6.7	Lung fibroblast TNF alpha + IL-1 beta	13.4
PBMC PWM	9.2	Lung fibroblast IL-4	21.8

PBMC PHA-L	10.0	Lung fibroblast IL-9	38.4
Ramos (B cell) none	29.7	Lung fibroblast IL-13	23.0
Ramos (B cell) ionomycin	29.3	Lung fibroblast IFN gamma	31.0
B lymphocytes PWM	9.3	Dermal fibroblast CCD1070 rest	23.7
B lymphocytes CD40L and IL-4	11.7	Dermal fibroblast CCD1070 TNF alpha	25.3
EOL-1 dbcAMP	15.0	Dermal fibroblast CCD1070 IL-1 beta	21.8
EOL-1 dbcAMP PMA/ionomycin	6.9	Dermal fibroblast IFN gamma	28.5
Dendritic cells none	40.3	Dermal fibroblast IL-4	78.5
Dendritic cells LPS	43.5	Dermal Fibroblasts rest	21.3
Dendritic cells anti-CD40	41.5	Neutrophils TNFa+LPS	6.9
Monocytes rest	22.1	Neutrophils rest	12.2
Monocytes LPS	46.3	Colon	0.0
Macrophages rest	36.6	Lung	20.4
Macrophages LPS	30.6	Thymus	24.1
HUVEC none	43.5	Kidney	61.6
HUVEC starved	53.6		

CNS_neurodegeneration_v1.0 Summary: Ag4926 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.5 Summary: Ag4926 Highest expression of this gene is seen in a colon cancer cell line (CT=26.7). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful

in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

In addition, this gene is expressed at much higher levels in fetal lung and liver tissue (CTs=28.5) when compared to expression in the adult counterparts (CTs=32.5). Thus, expression of this gene may be used to differentiate between the fetal and adult source of these tissues.

Panel 4.1D Summary: Ag4926 Highest expression of this gene is seen in HPAECst treated with TNF-a/IL1-b. This gene is also expressed at moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

D. NOV5 CG143198-01: Nuclear protein-like protein.

Expression of gene CG143198-01 was assessed using the primer-probe set Ag7138, described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB and DC.

Table DA. Probe Name Ag7138

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-caagactctgaagcagcagg-3'	20	1173	76
Probe	TET-5'-ttcgaaatttagatgctcagtatgaaatg-3'-TAMRA	29	1202	77
Reverse	5'-gtgtgggtttgtgatcttgc-3'	20	1231	78

Table DB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag7138, Run 283829331	issue Name	Rel. Exp.(%) Ag7138, Run 283829331
AD 1 Hippo	10.7	Control (Path) 3 Temporal Ctx	5.8
AD 2 Hippo	24.1	Control (Path) 4 Temporal Ctx	25.0
AD 3 Hippo	9.5	AD 1 Occipital Ctx	18.3
AD 4 Hippo	10.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	97.9	AD 3 Occipital Ctx	8.8
AD 6 Hippo	49.7	AD 4 Occipital Ctx	11.0
Control 2 Hippo	21.2	AD 5 Occipital Ctx	47.6
Control 4 Hippo	11.8	AD 6 Occipital Ctx	34.4
Control (Path) 3 Hippo	6.9	Control 1 Occipital Ctx	5.8
AD 1 Temporal Ctx	24.7	Control 2 Occipital Ctx	27.4
AD 2 Temporal Ctx	27.9	Control 3 Occipital Ctx	22.7
AD 3 Temporal Ctx	10.1	Control 4 Occipital Ctx	7.8
AD 4 Temporal Ctx	17.6	Control (Path) 1 Occipital Ctx	60.3
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	13.4
AD 5 Sup Temporal Ctx	55.1	Control (Path) 3 Occipital Ctx	3.9
AD 6 Inf Temporal Ctx	54.7	Control (Path) 4 Occipital Ctx	15.4
AD 6 Sup Temporal Ctx	52.5	Control 1 Parietal Ctx	7.3
Control 1 Temporal Ctx	6.7	Control 2 Parietal Ctx	56.3
Control 2 Temporal Ctx	25.0	Control 3 Parietal Ctx	13.3
Control 3 Temporal Ctx	16.5	Control (Path) 1 Parietal Ctx	53.2
Control 3 Temporal Ctx	9.5	Control (Path) 2 Parietal Ctx	22.2
Control (Path) 1 Temporal Ctx	45.7	Control (Path) 3 Parietal Ctx	7.2
Control (Path) 2 Temporal Ctx	31.9	Control (Path) 4 Parietal Ctx	27.0

Table DC. Panel 4.1D

Tissue Name	Rel. Exp.() Ag7138, Run 283838272	Tissue Name	Rel. Exp.(%) Ag7138, Run 283838272
Secondary Th1 act	55.9	HUVEC IL-1beta	30.4
Secondary Th2 act	**100.0**	HUVEC IFN gamma	27.5
Secondary Tr1 act	40.1	HUVEC TNF alpha + IFN gamma	11.7
Secondary Th1 rest	6.0	HUVEC TNF alpha + IL4	11.3
Secondary Th2 rest	8.4	HUVEC IL-11	10.4
Secondary Tr1 rest	7.6	Lung Microvascular EC none	38.7
Primary Th1 act	11.3	Lung Microvascular EC TNFalpha + IL-1beta	9.3

Primary Th2 act	53.6	Microvascular Dermal EC none	6.2
Primary Tr1 act	37.1	Microvascular Dermal EC TNFalpha + IL-1beta	9.7
Primary Th1 rest	6.3	Bronchial epithelium TNFalpha + IL1beta	9.5
Primary Th2 rest	4.1	Small airway epithelium none	6.1
Primary Tr1 rest	1.8	Small airway epithelium TNFalpha + IL-1beta	18.6
CD45RA CD4 lymphocyte act	40.9	Coronary artery SMC rest	12.9
CD45RO CD4 lymphocyte act	49.0	Coronary artery SMC TNFalpha + IL-1beta	17.3
CD8 lymphocyte act	17.6	Astrocytes rest	11.6
Secondary CD8 lymphocyte rest	8.2	Astrocytes TNFalpha + IL-1beta	5.6
Secondary CD8 lymphocyte act	12.5	KU-812 (Basophil) rest	27.4
CD4 lymphocyte none	7.3	KU-812 (Basophil) PMA/ionomycin	38.2
2ry Th1/Th2/Tr1_anti-CD95 CH11	10.7	CCD1106 (Keratinocytes) none	26.1
LAK cells rest	13.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.8
LAK cells IL-2	20.9	Liver cirrhosis	13.6
LAK cells IL-2+IL-12	2.8	NCI-H292 none	30.8
LAK cells IL-2+IFN gamma	10.5	NCI-H292 IL-4	30.6
LAK cells IL-2+ IL-18	7.5	NCI-H292 IL-9	46.0
LAK cells PMA/ionomycin	51.8	NCI-H292 IL-13	45.4
NK Cells IL-2 rest	45.7	NCI-H292 IFN gamma	24.7
Two Way MLR 3 day	25.5	HPAEC none	9.5
Two Way MLR 5 day	15.0	HPAEC TNF alpha + IL-1 beta	27.7
Two Way MLR 7 day	12.3	Lung fibroblast none	22.4
PBMC rest	5.6	Lung fibroblast TNF alpha + IL-1 beta	19.5
PBMC PWM	13.9	Lung fibroblast IL-4	12.2
PBMC PHA-L	22.2	Lung fibroblast IL-9	14.0
Ramos (B cell) none	12.2	Lung fibroblast IL-13	13.6
Ramos (B cell) ionomycin	40.9	Lung fibroblast IFN gamma	32.5
B lymphocytes PWM	25.0	Dermal fibroblast CCD1070 rest	37.1
B lymphocytes CD40L and IL-4	44.1	Dermal fibroblast CCD1070 TNF alpha	63.3
EOL-1 dbcAMP	53.2	Dermal fibroblast CCD1070 IL-1 beta	28.5
EOL-1 dbcAMP PMA/ionomycin	42.3	Dermal fibroblast IFN gamma	22.8
Dendritic cells none	15.0	Dermal fibroblast IL-4	32.3
Dendritic cells LPS	8.5	Dermal Fibroblasts rest	20.4
Dendritic cells anti-CD40	6.5	Neutrophils TNFa+LPS	6.3
Monocytes rest	8.0	Neutrophils rest	22.5

Monocytes LPS	38.2	Colon	4.0
Macrophages rest	8.7	Lung	2.1
Macrophages LPS	10.2	Thymus	9.8
HUVEC none	23.3	Kidney	28.5
HUVEC starved	31.4		

CNS_neurodegeneration_v1.0 Summary: Ag7138 This panel shows expression of this gene at moderate levels in the brain in an independent group of individuals. This gene appears to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

Panel 4.1D Summary: Ag7138 Highest expression of this gene is seen in chronically activated Th2 cells (CT=30.4). This gene is also expressed at moderate to low levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

E. NOV8 CG145988-01: Phosphatidylethanolamine-binding protein-like protein.

Expression of gene CG145988-01 was assessed using the primer-probe set Ag5948, described in Table EA.

Table EA. Probe Name Ag5948

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-atgcgggtgaacctcagac-3'	19	145	79
Probe	TET-5'-tggagccttcaggaagtggacgaa-3'-TAMRA	24	178	80
Reverse	5'-acccaggtgtagagtttcccta-3'	22	320	81

F. NOV9 CG146452-01: LRR domain containing protein.

Expression of gene CG146452-01 was assessed using the primer-probe set Ag7055, described in Table FA. Results of the RTQ-PCR runs are shown in Table FB.

Table FA. Probe Name Ag7055

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-gctgaggtagagctgctttagac-3'	23	507	82
Probe	TET-5'-tctactccaagcgcttcgccgtcttc-3'-TAMRA	26	552	83
Reverse	5'-ccttcatgctgcacctcat-3'	19	595	84

Table FB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag7055, Run 282273880	issue Name	Rel. Exp.(%) Ag7055, Run 282273880
Adipose	0.3	Renal ca. TK-10	33.4
Melanoma* Hs688(A).T	7.7	Bladder	9.0
Melanoma* Hs688(B).T	5.3	Gastric ca. (liver met.) NCI-N87	29.7
Melanoma* M14	8.1	Gastric ca. KATO III	27.2
Melanoma* LOXIMVI	0.7	Colon ca. SW-948	14.7
Melanoma* SK-MEL-5	1.1	Colon ca. SW480	32.5
Squamous cell carcinoma SCC-4	12.1	Colon ca.* (SW480 met) SW620	17.1
Testis Pool	1.1	Colon ca. HT29	14.5
Prostate ca.* (bone met) PC-3	25.9	Colon ca. HCT-116	47.6
Prostate Pool	3.0	Colon ca. CaCo-2	16.3
Placenta	13.9	Colon cancer tissue	12.6
Uterus Pool	0.4	Colon ca. SW1116	15.3
Ovarian ca. OVCAR-3	22.5	Colon ca. Colo-205	8.4
Ovarian ca. SK-OV-3	45.7	Colon ca. SW-48	11.0
Ovarian ca. OVCAR-4	22.8	Colon Pool	1.5
Ovarian ca. OVCAR-5	33.2	Small Intestine Pool	1.7
Ovarian ca. IGROV-1	33.0	Stomach Pool	2.3
Ovarian ca. OVCAR-8	53.2	Bone Marrow Pool	0.4
Ovary	1.3	Fetal Heart	0.5
Breast ca. MCF-7	32.8	Heart Pool	0.4
Breast ca. MDA-MB-231	31.6	Lymph Node Pool	1.9
Breast ca. BT 549	8.1	Fetal Skeletal Muscle	1.1
Breast ca. T47D	8.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	11.3	Spleen Pool	0.0
Breast Pool	1.8	Thymus Pool	4.4
Trachea	7.0	CNS cancer (glio/astro) U87-MG	100.0

Lung	0.4	CNS cancer (glio/astro) U-118-MG	10.2
Fetal Lung	3.1	CNS cancer (neuro;met) SK-N-AS	2.0
Lung ca. NCI-N417	1.4	CNS cancer (astro) SF-539	17.6
Lung ca. LX-1	11.2	CNS cancer (astro) SNB-75	40.9
Lung ca. NCI-H146	1.8	CNS cancer (glio) SNB-19	23.0
Lung ca. SHP-77	3.0	CNS cancer (glio) SF-295	20.4
Lung ca. A549	11.5	Brain (Amygdala) Pool	0.8
Lung ca. NCI-H526	6.2	Brain (cerebellum)	0.5
Lung ca. NCI-H23	20.9	Brain (fetal)	0.0
Lung ca. NCI-H460	10.7	Brain (Hippocampus) Pool	0.7
Lung ca. HOP-62	6.1	Cerebral Cortex Pool	0.6
Lung ca. NCI-H522	10.4	Brain (Substantia nigra) Pool	0.0
Liver	1.7	Brain (Thalamus) Pool	1.2
Fetal Liver	4.7	Brain (whole)	0.2
Liver ca. HepG2	11.4	Spinal Cord Pool	0.7
Kidney Pool	2.6	Adrenal Gland	5.3
Fetal Kidney	2.1	Pituitary gland Pool	1.7
Renal ca. 786-0	21.3	Salivary Gland	9.2
Renal ca. A498	20.3	Thyroid (female)	3.0
Renal ca. ACHN	16.3	Pancreatic ca. CAPAN2	9.2
Renal ca. UO-31	24.3	Pancreas Pool	7.7

General_screening_panel_v1.6 Summary: Ag7055 Highest expression of this gene is seen in a brain cancer cell line (CT=30). This gene is widely expressed in the cancer cell lines on this panel, with moderate to low expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at low but significant levels adrenal gland, pancreas, thyroid, and fetal liver. This expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

G. NOV10 CG146731-01: Membrane Binding protein-like protein.

Expression of gene CG146731-01 was assessed using the primer-probe set Ag6046, described in Table GA. Results of the RTQ-PCR runs are shown in Tables GB and GC.

Table GA. Probe Name Ag6046

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-catgaaccagccagagtctg-3'	20	30	85
Probe	TET-5'-gatcctgaacccctgtgtgcagtgt-3'-TAMRA	25	55	86
Reverse	5'-gaagtggttttcctccaagg-3'	20	95	87

Table GB. General screening panel v1.5

5

Tissue Name	Rel. Exp.(%) Ag6046, Run 228783203	issue Name	Rel. Exp.(%) Ag6046, Run 228783203
Adipose	9.4	Renal ca. TK-10	10.9
Melanoma* Hs688(A).T	0.0	Bladder	36.1
Melanoma* Hs688(B).T	0.1	Gastric ca. (liver met.) NCI-N87	45.1
Melanoma* M14	0.0	Gastric ca. KATO III	9.2
Melanoma* LOXIMVI	0.1	Colon ca. SW-948	9.6
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.7
Squamous cell carcinoma SCC-4	2.3	Colon ca.* (SW480 met) SW620	1.0
Testis Pool	3.7	Colon ca. HT29	29.5
Prostate ca.* (bone met) PC-3	1.2	Colon ca. HCT-116	4.8
Prostate Pool	17.0	Colon ca. CaCo-2	39.0
Placenta	51.1	Colon cancer tissue	29.3
Uterus Pool	18.7	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	45.4	Colon ca. Colo-205	17.2
Ovarian ca. SK-OV-3	0.3	Colon ca. SW-48	31.9
Ovarian ca. OVCAR-4	2.4	Colon Pool	15.8
Ovarian ca. OVCAR-5	16.7	Small Intestine Pool	11.4
Ovarian ca. IGROV-1	0.4	Stomach Pool	13.0
Ovarian ca. OVCAR-8	0.6	Bone Marrow Pool	7.6
Ovary	4.8	Fetal Heart	1.0
Breast ca. MCF-7	47.3	Heart Pool	8.8
Breast ca. MDA-MB-231	4.6	Lymph Node Pool	15.7
Breast ca. BT 549	1.1	Fetal Skeletal Muscle	1.3
Breast ca. T47D	10.7	Skeletal Muscle Pool	17.6
Breast ca. MDA-N	0.0	Spleen Pool	1.1
Breast Pool	8.3	Thymus Pool	17.0
Trachea	51.1	CNS cancer (glio/astro) U87-MG	0.1
Lung	1.1	CNS cancer (glio/astro) U-118-MG	0.6
Fetal Lung	10.5	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.1	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	5.6	CNS cancer (astro) SNB-75	0.2
Lung ca. NCI-H146	100.0	CNS cancer (glio) SNB-19	1.5

Lung ca. SHP-77	3.3	CNS cancer (glio) SF-295	0.3
Lung ca. A549	0.1	Brain (Amygdala) Pool	0.3
Lung ca. NCI-H526	8.5	Brain (cerebellum)	0.7
Lung ca. NCI-H23	1.7	Brain (fetal)	0.5
Lung ca. NCI-H460	3.5	Brain (Hippocampus) Pool	0.3
Lung ca. HOP-62	0.2	Cerebral Cortex Pool	0.2
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.2
Liver	2.7	Brain (Thalamus) Pool	0.3
Fetal Liver	10.2	Brain (whole)	1.4
Liver ca. HepG2	21.2	Spinal Cord Pool	0.3
Kidney Pool	23.0	Adrenal Gland	3.8
Fetal Kidney	36.1	Pituitary gland Pool	10.6
Renal ca. 786-0	0.0	Salivary Gland	71.7
Renal ca. A498	0.2	Thyroid (female)	26.6
Renal ca. ACHN	0.2	Pancreatic ca. CAPAN2	27.4
Renal ca. UO-31	0.3	Pancreas Pool	25.5

Table GC. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag6046, Run 225160585	Tissue Name	Rel. Exp.(%) Ag6046, Run 225160585
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.7	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.3	HUVEC IL-11	1.6
Secondary Tr1 rest	0.0	Lung Microvascular EC none	1.4
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.3
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	4.6
Primary Th2 rest	0.0	Small airway epithelium none	6.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	2.2
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.1	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	35.8

CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	31.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.5
LAK cells rest	0.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.6
LAK cells IL-2	0.0	Liver cirrhosis	15.6
LAK cells IL-2+IL-12	0.3	NCI-H292 none	13.9
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	9.6
LAK cells IL-2+ IL-18	0.8	NCI-H292 IL-9	18.6
LAK cells PMA/ionomycin	1.4	NCI-H292 IL-13	8.6
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	14.9
Two Way MLR 3 day	0.0	HPAEC none	0.3
Two Way MLR 5 day	1.0	HPAEC TNF alpha + IL-1 beta	0.2
Two Way MLR 7 day	0.9	Lung fibroblast none	2.1
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	1.8
PBMC PWM	0.3	Lung fibroblast IL-4	2.1
PBMC PHA-L	0.3	Lung fibroblast IL-9	2.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	1.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.4
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.2
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.2
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	1.4
Dendritic cells none	0.4	Dermal fibroblast IL-4	0.3
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	1.3
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.3
Monocytes rest	0.0	Neutrophils rest	0.8
Monocytes LPS	10.5	Colon	12.8
Macrophages rest	0.6	Lung	11.1
Macrophages LPS	1.3	Thymus	31.2
HUVEC none	0.6	Kidney	**100.0**
HUVEC starved	0.0		

General_screening_panel_v1.5 Summary: Ag6046 Highest expression of this gene is seen in a lung cancer cell line (CT=28.7). Moderate levels of expression are also seen in ovarian and breast cancer cell lines relative to the expression in the normal tissue samples. Thus, therapeutic modulation of the expression or function of this gene may be effective in the treatment of these cancers.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, heart, skeletal muscle, and adult and fetal liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is expressed at much higher levels in heart and skeletal muscle (CTs=31-32) when compared to expression in their fetal counterpart (CTs=35). Conversely, expression of this gene is much higher in fetal lung tissue (CT=32) when compared to expression in the adult counterpart (CT=35). Thus, expression of this gene may be used to differentiate between the fetal and adult sources of these tissues. In addition, the relative overexpression of this gene in fetal lung tissue suggests that the protein product may enhance lung growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

Panel 4.1D Summary: Ag6046 Highest expression of this gene is seen in the kidney (CT=30), with low to moderate expression seen in LPS treated monocytes, untreated small airway epithelium, TNF-a/IL1-b treated bronchial epithelium, liver cirrhosis, treated and untreated samples from the KU-812 basophil cell line and the NCI-H292 mucoepidermoid cell line, and normal colon, lung and thymus. Thus, expression of this gene could be used to differentiate the kidney derived sample from other samples on this panel and as a marker of kidney tissue. In addition, therapeutic targeting of the expression or function of this gene may modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

H. NOV12 CG147246-01: Actin-Binding Protein Frabin-Alpha -like protein.

Expression of gene CG147246-01 was assessed using the primer-probe set Ag6047, described in Table HA. Results of the RTQ-PCR runs are shown in Tables HB, HC and HD.

Table HA. Probe Name Ag6047

Primers	Sequence	Length	Start Position	SEQ ID No

Forward	5'-tcggaacacttcagcacaa-3'	19	1305	88
Probe	TET-5'-ccttttcttattcaacaacatgttgctg-3'-TAMRA	28	1332	89
Reverse	5'-ctggatttgggcacacagta-3'	20	1360	90

Table HB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag6047, Run 225249597	issue Name	Rel. Exp.(%) Ag6047, Run 225249597
AD 1 Hippo	16.8	Control (Path) 3 Temporal Ctx	11.3
AD 2 Hippo	41.2	Control (Path) 4 Temporal Ctx	39.5
AD 3 Hippo	11.7	AD 1 Occipital Ctx	28.1
AD 4 Hippo	11.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	59.9	AD 3 Occipital Ctx	13.8
AD 6 Hippo	73.7	AD 4 Occipital Ctx	31.0
Control 2 Hippo	33.7	AD 5 Occipital Ctx	46.7
Control 4 Hippo	26.4	AD 6 Occipital Ctx	34.9
Control (Path) 3 Hippo	11.1	Control 1 Occipital Ctx	6.7
AD 1 Temporal Ctx	25.7	Control 2 Occipital Ctx	39.2
AD 2 Temporal Ctx	38.4	Control 3 Occipital Ctx	19.6
AD 3 Temporal Ctx	9.0	Control 4 Occipital Ctx	14.8
AD 4 Temporal Ctx	28.1	Control (Path) 1 Occipital Ctx	**100.0**
AD 5 Inf Temporal Ctx	72.7	Control (Path) 2 Occipital Ctx	13.0
AD 5 SupTemporal Ctx	30.4	Control (Path) 3 Occipital Ctx	5.5
AD 6 Inf Temporal Ctx	61.1	Control (Path) 4 Occipital Ctx	18.6
AD 6 Sup Temporal Ctx	24.3	Control 1 Parietal Ctx	12.1
Control 1 Temporal Ctx	7.8	Control 2 Parietal Ctx	35.1
Control 2 Temporal Ctx	27.9	Control 3 Parietal Ctx	20.2
Control 3 Temporal Ctx	14.8	Control (Path) 1 Parietal Ctx	63.3
Control 4 Temporal Ctx	14.8	Control (Path) 2 Parietal Ctx	29.9
Control (Path) 1 Temporal Ctx	44.1	Control (Path) 3 Parietal Ctx	11.2
Control (Path) 2 Temporal Ctx	25.2	Control (Path) 4 Parietal Ctx	47.0

Table HC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag6047, Run 228783233	issue Name	Rel. Exp.(%) Ag6047, Run 228783233
Adipose	49.7	Renal ca. TK-10	31.2
Melanoma* Hs688(A).T	2.4	Bladder	71.2
Melanoma* Hs688(B).T	8.5	Gastric ca. (liver met.) NCI-N87	**100.0**
Melanoma* M14	14.9	Gastric ca. KATO III	63.7

Melanoma* LOXIMVI	0.8	Colon ca. SW-948	9.1
Melanoma* SK-MEL-5	63.3	Colon ca. SW480	47.3
Squamous cell carcinoma SCC-4	3.7	Colon ca.* (SW480 met) SW620	14.4
Testis Pool	36.6	Colon ca. HT29	24.1
Prostate ca.* (bone met) PC-3	3.3	Colon ca. HCT-116	20.6
Prostate Pool	27.7	Colon ca. CaCo-2	50.7
Placenta	3.7	Colon cancer tissue	52.9
Uterus Pool	21.8	Colon ca. SW1116	1.5
Ovarian ca. OVCAR-3	43.5	Colon ca. Colo-205	6.6
Ovarian ca. SK-OV-3	80.7	Colon ca. SW-48	9.5
Ovarian ca. OVCAR-4	11.3	Colon Pool	14.0
Ovarian ca. OVCAR-5	29.5	Small Intestine Pool	7.5
Ovarian ca. IGROV-1	26.8	Stomach Pool	15.1
Ovarian ca. OVCAR-8	4.0	Bone Marrow Pool	8.0
Ovary	23.3	Fetal Heart	22.1
Breast ca. MCF-7	14.4	Heart Pool	10.7
Breast ca. MDA-MB-231	7.5	Lymph Node Pool	13.2
Breast ca. BT 549	0.2	Fetal Skeletal Muscle	24.0
Breast ca. T47D	11.3	Skeletal Muscle Pool	66.9
Breast ca. MDA-N	10.8	Spleen Pool	11.2
Breast Pool	14.6	Thymus Pool	13.0
Trachea	24.7	CNS cancer (glio/astro) U87-MG	0.0
Lung	18.9	CNS cancer (glio/astro) U-118-MG	21.2
Fetal Lung	73.2	CNS cancer (neuro;met) SK-N-AS	10.4
Lung ca. NCI-N417	9.9	CNS cancer (astro) SF-539	0.4
Lung ca. LX-1	28.7	CNS cancer (astro) SNB-75	15.0
Lung ca. NCI-H146	23.7	CNS cancer (glio) SNB-19	42.0
Lung ca. SHP-77	19.8	CNS cancer (glio) SF-295	10.8
Lung ca. A549	9.5	Brain (Amygdala) Pool	28.7
Lung ca. NCI-H526	9.9	Brain (cerebellum)	19.9
Lung ca. NCI-H23	7.2	Brain (fetal)	65.5
Lung ca. NCI-H460	25.7	Brain (Hippocampus) Pool	32.3
Lung ca. HOP-62	17.7	Cerebral Cortex Pool	35.4
Lung ca. NCI-H522	14.3	Brain (Substantia nigra) Pool	25.0
Liver	1.7	Brain (Thalamus) Pool	26.8
Fetal Liver	11.4	Brain (whole)	7.2
Liver ca. HepG2	16.7	Spinal Cord Pool	26.4
Kidney Pool	17.6	Adrenal Gland	7.7
Fetal Kidney	24.1	Pituitary gland Pool	14.2
Renal ca. 786-0	40.1	Salivary Gland	2.4
Renal ca. A498	7.8	Thyroid (female)	6.4
Renal ca. ACHN	49.0	Pancreatic ca. CAPAN2	50.0
Renal ca. UO-31	29.9	Pancreas Pool	20.9

Table HD. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag6047, Run 225160587	Tissue Name	Rel. Exp.(%) Ag6047, Run 225160587
Secondary Th1 act	0.0	HUVEC IL-1beta	36.3
Secondary Th2 act	0.4	HUVEC IFN gamma	26.4
Secondary Tr1 act	0.4	HUVEC TNF alpha + IFN gamma	11.7
Secondary Th1 rest	1.0	HUVEC TNF alpha + IL4	24.1
Secondary Th2 rest	0.0	HUVEC IL-11	23.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	26.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	5.5
Primary Th2 act	0.0	Microvascular Dermal EC none	22.1
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	2.9
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	20.3
Primary Th2 rest	0.2	Small airway epithelium none	12.2
Primary Tr1 rest	0.6	Small airway epithelium TNFalpha + IL-1beta	23.5
CD45RA CD4 lymphocyte act	1.7	Coronary artery SMC rest	5.0
CD45RO CD4 lymphocyte act	0.4	Coronary artery SMC TNFalpha + IL-1beta	3.9
CD8 lymphocyte act	0.6	Astrocytes rest	5.0
Secondary CD8 lymphocyte rest	0.9	Astrocytes TNFalpha + IL-1beta	1.4
Secondary CD8 lymphocyte act	1.9	KU-812 (Basophil) rest	5.3
CD4 lymphocyte none	0.9	KU-812 (Basophil) PMA/ionomycin	3.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.3	CCD1106 (Keratinocytes) none	18.0
LAK cells rest	27.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	13.5
LAK cells IL-2	0.0	Liver cirrhosis	26.4
LAK cells IL-2+IL-12	0.2	NCI-H292 none	14.3
LAK cells IL-2+IFN gamma	0.8	NCI-H292 IL-4	5.3
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	25.3
LAK cells PMA/ionomycin	13.3	NCI-H292 IL-13	22.5
NK Cells IL-2 rest	1.1	NCI-H292 IFN gamma	15.9
Two Way MLR 3 day	13.0	HPAEC none	12.8
Two Way MLR 5 day	9.7	HPAEC TNF alpha + IL-1 beta	8.8
Two Way MLR 7 day	2.1	Lung fibroblast none	13.2
PBMC rest	10.1	Lung fibroblast TNF alpha + IL-1 beta	9.3
PBMC PWM	0.0	Lung fibroblast IL-4	8.1

PBMC PHA-L	1.4	Lung fibroblast IL-9	14.9
Ramos (B cell) none	25.5	Lung fibroblast IL-13	5.2
Ramos (B cell) ionomycin	26.1	Lung fibroblast IFN gamma	10.2
B lymphocytes PWM	0.2	Dermal fibroblast CCD1070 rest	6.3
B lymphocytes CD40L and IL-4	0.7	Dermal fibroblast CCD1070 TNF alpha	6.7
EOL-1 dbcAMP	0.9	Dermal fibroblast CCD1070 IL-1 beta	7.9
EOL-1 dbcAMP PMA/ionomycin	0.2	Dermal fibroblast IFN gamma	7.1
Dendritic cells none	55.9	Dermal fibroblast IL-4	7.6
Dendritic cells LPS	7.5	Dermal Fibroblasts rest	7.7
Dendritic cells anti-CD40	54.7	Neutrophils TNFa+LPS	**100.0**
Monocytes rest	59.0	Neutrophils rest	38.7
Monocytes LPS	10.4	Colon	12.6
Macrophages rest	75.3	Lung	19.9
Macrophages LPS	21.2	Thymus	10.6
HUVEC none	24.3	Kidney	33.7
HUVEC starved	46.3		

CNS_neurodegeneration_v1.0 Summary: Ag6047 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Please see Panel 1.5 for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.5 Summary: Ag6047 Highest expression of this gene is seen in a gastric cancer cell line (CT=27.4). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of these cancers.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful

in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag6047 Highest expression of this gene is seen in TNFa/LPS treated neutrophils (CT=28.4). Moderate levels of expression are also seen in many cell types, including LAK cells, monocytes, macrophages, dendritic cells, keratinocytes, lung epithelium, lung and skin endothelium, lung and dermal fibroblasts, HPAECs, HUVECs, and treated and untreated samples from the NCI-H292 mucoepidermoid cell line. The neutrophil expression is reduced in resting neutrophils, suggesting that the protein encoded by this gene is produced by activated neutrophils but not by resting neutrophils. Thus, expression of this gene could be used to differentiate between resting and activated neutrophils. In addition, modulation of the expression or function of this gene may be useful in the treatment of symptoms in patients with Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

I. NOV13 CG147651-01: Actin Related Protein 2/3 Complex protein-like protein.

Expression of gene CG147651-01 was assessed using the primer-probe set Ag5949, described in Table IA. Results of the RTQ-PCR runs are shown in Tables IB and IC.

Table IA. Probe Name Ag5949

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-ccagattgccctcagtcctg-3'	20	120	91
Probe	TET-5'-catgggagccagcgagtgaaagct-3'-TAMRA	24	169	92
Reverse	5'-tgatatgtccattgtgctcactgag-3'	25	199	93

Table IB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag5949, Run 247854400	issue Name	Rel. Exp.(%) Ag5949, Run 247854400
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	0.0
AD 2 Hippo	20.3	Control (Path) 4 Temporal Ctx	0.0
AD 3 Hippo	0.0	AD 1 Occipital Ctx	0.0
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	43.5	AD 3 Occipital Ctx	0.0
AD 6 Hippo	45.4	AD 4 Occipital Ctx	13.2

Control 2 Hippo	0.0	AD 5 Occipital Ctx	0.0
Control 4 Hippo	13.6	AD 6 Occipital Ctx	15.5
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	24.5	Control 2 Occipital Ctx	25.3
AD 2 Temporal Ctx	0.0	Control 3 Occipital Ctx	0.0
AD 3 Temporal Ctx	12.9	Control 4 Occipital Ctx	0.0
AD 4 Temporal Ctx	9.1	Control (Path) 1 Occipital Ctx	38.2
AD 5 Inf Temporal Ctx	**100.0**	Control (Path) 2 Occipital Ctx	0.0
AD 5 SupTemporal Ctx	42.3	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	0.0	Control (Path) 4 Occipital Ctx	43.5
AD 6 Sup Temporal Ctx	63.3	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	13.8
Control 2 Temporal Ctx	0.0	Control 3 Parietal Ctx	0.0
Control 3 Temporal Ctx	0.0	Control (Path) 1 Parietal Ctx	40.3
Control 4 Temporal Ctx	0.0	Control (Path) 2 Parietal Ctx	0.0
Control (Path) 1 Temporal Ctx	17.4	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	45.4	Control (Path) 4 Parietal Ctx	15.1

Table IC. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag5949, Run 247850160	Tissue Name	Rel. Exp.(%) Ag5949, Run 247850160
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	39.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	42.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	96.6
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	63.3
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	24.1
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0

Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	**100.0**
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	29.3
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	17.7
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	16.6
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	38.7
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	77.9
Two Way MLR 7 day	0.0	Lung fibroblast none	23.8
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	24.5
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	74.2
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	21.9
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	27.9
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	12.2
Dendritic cells none	0.0	Dermal fibroblast IL-4	41.2
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	35.6
HUVEC starved	27.9		

CNS_neurodegeneration_v1.0 Summary: Ag5949 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile shows this gene to be expressed at low levels in the brain. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders,

such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag5949 Expression in this panel is restricted to untreated lung microvascular endothelial cells, IFN gamma treated lung fibroblasts, untreated keratinocytes, untreated coronary artery smooth muscle cells, and TNF-a/IL1-b treated HPAECs. Thus, this gene product may be involved in inflammatory conditions of the lung, including asthma, allergy, emphysema, and COPD.

J. NOV14 CG149303-01: Hepatocellular Carcinoma Autoantigen - Like protein.

Expression of gene CG149303-01 was assessed using the primer-probe set Ag5631, described in Table JA. Results of the RTQ-PCR runs are shown in Tables JB, JC and JD.

Table JA. Probe Name Ag5631

Primers	Sequence	Length	Start Position	SEQ ID No
Forward	5'-tgccagtgctgagatagagatt-3'	22	1006	94
Probe	TET-5'-attttcaaaggcctcacgcagcttct-3'-TAMRA	26	1032	95
Reverse	5'-ccgaagtgggtattaacagtca-3'	22	1065	96

Table JB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag5631, Run 246956912	issue Name	Rel. Exp.(%) Ag5631, Run 246956912
AD 1 Hippo	7.5	Control (Path) 3 Temporal Ctx	6.7
AD 2 Hippo	22.8	Control (Path) 4 Temporal Ctx	49.7
AD 3 Hippo	0.0	AD 1 Occipital Ctx	2.7
AD 4 Hippo	24.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	31.0	AD 3 Occipital Ctx	0.0
AD 6 Hippo	37.6	AD 4 Occipital Ctx	35.1
Control 2 Hippo	62.9	AD 5 Occipital Ctx	27.2
Control 4 Hippo	19.3	AD 6 Occipital Ctx	9.0
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	12.9
AD 1 Temporal Ctx	26.2	Control 2 Occipital Ctx	63.7
AD 2 Temporal Ctx	40.9	Control 3 Occipital Ctx	86.5
AD 3 Temporal Ctx	7.0	Control 4 Occipital Ctx	21.6
AD 4 Temporal Ctx	31.6	Control (Path) 1 Occipital Ctx	42.0
AD 5 Inf Temporal Ctx	49.7	Control (Path) 2 Occipital Ctx	31.6
AD 5 Sup Temporal Ctx	25.0	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	46.3	Control (Path) 4 Occipital Ctx	18.7

AD 6 Sup Temporal Ctx	88.3	Control 1 Parietal Ctx	13.8
Control 1 Temporal Ctx	15.3	Control 2 Parietal Ctx	35.6
Control 2 Temporal Ctx	34.6	Control 3 Parietal Ctx	57.8
Control 3 Temporal Ctx	**100.0**	Control (Path) 1 Parietal Ctx	49.3
Control 3 Temporal Ctx	29.7	Control (Path) 2 Parietal Ctx	41.5
Control (Path) 1 Temporal Ctx	66.9	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	81.8	Control (Path) 4 Parietal Ctx	69.7

Table JC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag5631, Run 245240923	issue Name	Rel. Exp.(%) Ag5631, Run 245240923
Adipose	36.9	Renal ca. TK-10	4.6
Melanoma* Hs688(A).T	9.4	Bladder	38.7
Melanoma* Hs688(B).T	17.4	Gastric ca. (liver met.) NCI-N87	8.6
Melanoma* M14	20.2	Gastric ca. KATO III	16.6
Melanoma* LOXIMVI	57.0	Colon ca. SW-948	46.3
Melanoma* SK-MEL-5	88.9	Colon ca. SW480	3.7
Squamous cell carcinoma SCC-4	16.7	Colon ca.* (SW480 met) SW620	70.2
Testis Pool	1.3	Colon ca. HT29	40.6
Prostate ca.* (bone met) PC-3	19.2	Colon ca. HCT-116	19.5
Prostate Pool	58.6	Colon ca. CaCo-2	16.0
Placenta	8.2	Colon cancer tissue	29.9
Uterus Pool	11.4	Colon ca. SW1116	10.9
Ovarian ca. OVCAR-3	8.1	Colon ca. Colo-205	1.7
Ovarian ca. SK-OV-3	63.3	Colon ca. SW-48	9.9
Ovarian ca. OVCAR-4	11.7	Colon Pool	77.9
Ovarian ca. OVCAR-5	3.4	Small Intestine Pool	81.2
Ovarian ca. IGROV-1	**100.0**	Stomach Pool	8.6
Ovarian ca. OVCAR-8	47.0	Bone Marrow Pool	61.1
Ovary	20.3	Fetal Heart	76.8
Breast ca. MCF-7	6.9	Heart Pool	22.1
Breast ca. MDA-MB-231	14.1	Lymph Node Pool	11.1
Breast ca. BT 549	1.3	Fetal Skeletal Muscle	12.7
Breast ca. T47D	5.0	Skeletal Muscle Pool	28.9
Breast ca. MDA-N	70.7	Spleen Pool	75.8
Breast Pool	2.6	Thymus Pool	0.0
Trachea	37.1	CNS cancer (glio/astro) U87-MG	19.1
Lung	46.3	CNS cancer (glio/astro) U-118-MG	10.4
Fetal Lung	8.7	CNS cancer (neuro;met) SK-N-AS	67.8
Lung ca. NCI-N417	8.8	CNS cancer (astro) SF-539	10.2
Lung ca. LX-1	5.6	CNS cancer (astro) SNB-75	21.9

Lung ca. NCI-H146	14.6	CNS cancer (glio) SNB-19	9.0
Lung ca. SHP-77	3.3	CNS cancer (glio) SF-295	8.9
Lung ca. A549	6.3	Brain (Amygdala) Pool	66.9
Lung ca. NCI-H526	2.2	Brain (cerebellum)	32.1
Lung ca. NCI-H23	6.2	Brain (fetal)	23.3
Lung ca. NCI-H460	7.0	Brain (Hippocampus) Pool	12.3
Lung ca. HOP-62	6.8	Cerebral Cortex Pool	65.5
Lung ca. NCI-H522	5.0	Brain (Substantia nigra) Pool	15.2
Liver	4.3	Brain (Thalamus) Pool	4.5
Fetal Liver	7.6	Brain (whole)	10.4
Liver ca. HepG2	11.5	Spinal Cord Pool	81.8
Kidney Pool	10.3	Adrenal Gland	43.2
Fetal Kidney	52.1	Pituitary gland Pool	5.9
Renal ca. 786-0	20.9	Salivary Gland	37.1
Renal ca. A498	20.4	Thyroid (female)	1.8
Renal ca. ACHN	2.7	Pancreatic ca. CAPAN2	64.2
Renal ca. UO-31	4.3	Pancreas Pool	7.7

Table JD. Panel 4.1D

Tissue Name	Rel. Ep.(%) Ag5631, Run 246490837	Tissue Name	Rel. Exp.(%) Ag5631, Run 246490837
Secondary Th1 act	24.7	HUVEC IL-1beta	8.3
Secondary Th2 act	93.3	HUVEC IFN gamma	25.5
Secondary Tr1 act	5.4	HUVEC TNF alpha + IFN gamma	7.5
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	16.4
Secondary Tr1 rest	0.0	Lung Microvascular EC none	46.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.4
Primary Th2 act	34.6	Microvascular Dermal EC none	3.0
Primary Tr1 act	26.6	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	14.5
Primary Th2 rest	0.0	Small airway epithelium none	12.6
Primary Tr1 rest	3.4	Small airway epithelium TNFalpha + IL-1beta	23.2
CD45RA CD4 lymphocyte act	8.6	Coronary artery SMC rest	14.0
CD45RO CD4 lymphocyte act	88.3	Coronary artery SMC TNFalpha + IL-1beta	13.9
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	16.0	Astrocytes TNFalpha + IL-1beta	1.2

Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	31.4
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	60.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	9.8	CCD1106 (Keratinocytes) none	12.1
LAK cells rest	12.6	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	15.0
LAK cells IL-2	2.8	Liver cirrhosis	4.8
LAK cells IL-2+IL-12	0.0	NCI-H292 none	6.9
LAK cells IL-2+IFN gamma	4.0	NCI-H292 IL-4	8.2
LAK cells IL-2+ IL-18	3.4	NCI-H292 IL-9	35.1
LAK cells PMA/ionomycin	24.0	NCI-H292 IL-13	17.8
NK Cells IL-2 rest	56.6	NCI-H292 IFN gamma	10.5
Two Way MLR 3 day	7.9	HPAEC none	7.9
Two Way MLR 5 day	1.2	HPAEC TNF alpha + IL-1 beta	18.7
Two Way MLR 7 day	15.3	Lung fibroblast none	22.2
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	16.7
PBMC PWM	11.0	Lung fibroblast IL-4	9.2
PBMC PHA-L	7.4	Lung fibroblast IL-9	11.8
Ramos (B cell) none	3.0	Lung fibroblast IL-13	4.5
Ramos (B cell) ionomycin	22.8	Lung fibroblast IFN gamma	18.4
B lymphocytes PWM	11.8	Dermal fibroblast CCD1070 rest	29.5
B lymphocytes CD40L and IL-4	54.0	Dermal fibroblast CCD1070 TNF alpha	**100.0**
EOL-1 dbcAMP	29.5	Dermal fibroblast CCD1070 IL-1 beta	27.7
EOL-1 dbcAMP PMA/ionomycin	8.9	Dermal fibroblast IFN gamma	14.9
Dendritic cells none	2.6	Dermal fibroblast IL-4	57.4
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	12.7
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	14.2
Monocytes rest	3.6	Neutrophils rest	49.0
Monocytes LPS	52.1	Colon	0.0
Macrophages rest	5.1	Lung	0.0
Macrophages LPS	3.7	Thymus	0.0
HUVEC none	8.9	Kidney	23.0
HUVEC starved	13.0		

CNS_neurodegeneration_v1.0 Summary: Ag5631 Low expression of this gene is seen in temporal cortex of a control patient (CT=34.7). Therefore, therapeutic modulation of this gene may be useful in treatment of neurological disorders.

General_screening_panel_v1.5 Summary: Ag5631 Highest expression of this gene is detected in ovarian cancer IGROV-1 cell line (CT=32.8). Moderate to low levels of

expression of this gene is also seen in number of cancer cell lines derived from pancreatic, colon, ovarian, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, colon, ovarian, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in adipose, adrenal gland, skeletal muscle, fetal heart, and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at low levels in amygdala, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag5631 Highest expression of this gene is detected in TNF alpha treated dermal fibroblasts (CT=33.3). Low levels of expression of this gene is also seen in IL-4 activated dermal fibroblasts, resting neutrophils, basophils, lung microvascular endothelial cells, LPS activated monocytes, CD40L and IL-4 activated B lymphocytes, resting IL-2 treated NK cells, activated primary and secondary Th2 cells and activated memory T cells (CD45RO CD4 lymphocyte). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

K. NOV15 CG149312-01: Hematopoietic stem/progenitor cells protein MDS029 - Like protein.

Expression of gene CG149312-01 was assessed using the primer-probe set Ag5846, described in Table KA.

Table KA. Probe Name Ag5846

Primers		Length	Start Position	SEQ ID No
Forward	5'-atgggtctcacacaaaacaca-3'	21	260	97
Probe	TET-5'-tgtcccacgttgactccagtctcttc-3'-TAMRA	26	283	98
Reverse	5'-ctccagtctcttcgttgtgttt-3'	22	274	99

L. NOV16 CG150951-02: TRAP-delta protein-like protein.

Expression of gene CG150951-02 was assessed using the primer-probe set Ag6940, described in Table LA. Results of the RTQ-PCR runs are shown in Table LB. Please note that this sequence represents a full-length physical clone.

Table LA. Probe Name Ag6940

Primers		Length	Start Position	SEQ ID No
Forward	5'-gagatctccctgacatgcaa-3'	20	315	100
Probe	TET-5'-ctccaggacacctggaccctgttc-3'-TAMRA	24	335	101
Reverse	5'-ctcgtcgaagaatctaacctcatag-3'	25	389	102

Table LB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6940, Run 278700428	issue Name	Rel. Exp.(%) Ag6940, Run 278700428
Adipose	0.0	Renal ca. TK-10	2.6
Melanoma* Hs688(A).T	1.5	Bladder	3.1
Melanoma* Hs688(B).T	7.5	Gastric ca. (liver met.) NCI-N87	1.3
Melanoma* M14	6.3	Gastric ca. KATO III	29.9
Melanoma* LOXIMVI	9.7	Colon ca. SW-948	0.7
Melanoma* SK-MEL-5	0.8	Colon ca. SW480	33.4
Squamous cell carcinoma SCC-4	**100.0**	Colon ca.* (SW480 met) SW620	1.9
Testis Pool	0.0	Colon ca. HT29	6.2
Prostate ca.* (bone met) PC-3	7.0	Colon ca. HCT-116	9.7
Prostate Pool	0.0	Colon ca. CaCo-2	8.1
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	2.9	Colon ca. SW1116	0.8
Ovarian ca. OVCAR-3	10.6	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	9.7	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	7.3	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	6.6	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	2.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	14.1	Heart Pool	0.0
Breast ca. MDA-MB-231	8.2	Lymph Node Pool	1.2
Breast ca. BT 549	22.1	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	2.1
Breast ca. MDA-N	14.4	Spleen Pool	0.0

Breast Pool	0.0	Thymus Pool	4.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	18.2
Lung	0.0	CNS cancer (glio/astro) U-118-MG	8.4
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	19.5	CNS cancer (astro) SF-539	18.3
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	5.1
Lung ca. NCI-H146	9.9	CNS cancer (glio) SNB-19	3.0
Lung ca. SHP-77	14.6	CNS cancer (glio) SF-295	0.0
Lung ca. A549	4.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	6.8	Brain (cerebellum)	0.0
Lung ca. NCI-H23	80.1	Brain (fetal)	0.5
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	1.3	Adrenal Gland	0.0
Fetal Kidney	0.8	Pituitary gland Pool	0.0
Renal ca. 786-0	4.2	Salivary Gland	0.0
Renal ca. A498	0.9	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	5.7
Renal ca. UO-31	1.2	Pancreas Pool	1.6

General_screening_panel_v1.6 Summary: Ag6940 Highest expression of this gene is detected in squamous cell carcinoma SCC-4 cell line (CT=32). In addition, moderate to low levels of expression of this gene is seen mainly in number of cancer cell lines derived from lung, gastric, colon and brain cancers. Therefore, expression of this gene may be used as diagnostic marker to detect the presence of squamous cell carcinoma, lung, gastric, colon, and brain cancer. Furthermore, therapeutic modulation of this gene or its protein product may be useful in the treatment of these cancers.

M. NOV20 CG59323-02 and CG59323-03: TP53BP2: tumor protein p53-binding protein (ASSP) -like protein.

Expression of gene CG59323-02 and CG59323-03 was assessed using the primer-probe set Ag6327, described in Table MA. Results of the RTQ-PCR runs are shown in Tables MB, MC, MD and ME.

Table MA. Probe Name Ag6327

Primers		Length	Start Position	SEQ ID No
Forward	5'-atgaggttgatgacccaagc-3'	20	2897	103
Probe	TET-5'-tgcccaatgatgaaggcatcacg-3'-TAMRA	23	2918	104
Reverse	5'-acacacagcattgtgaagagc-3'	21	2941	105

Table MB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag6327, Run 259045579	issue Name	Rel. Exp.(%) Ag6327, Run 259045579
AD 1 Hippo	11.7	Control (Path) 3 Temporal Ctx	53.6
AD 2 Hippo	88.9	Control (Path) 4 Temporal Ctx	23.2
AD 3 Hippo	4.4	AD 1 Occipital Ctx	14.8
AD 4 Hippo	25.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	35.8	AD 3 Occipital Ctx	6.0
AD 6 Hippo	48.0	AD 4 Occipital Ctx	48.3
Control 2 Hippo	19.9	AD 5 Occipital Ctx	24.5
Control 4 Hippo	17.7	AD 6 Occipital Ctx	17.9
Control (Path) 3 Hippo	52.9	Control 1 Occipital Ctx	4.6
AD 1 Temporal Ctx	17.4	Control 2 Occipital Ctx	22.1
AD 2 Temporal Ctx	100.0	Control 3 Occipital Ctx	18.0
AD 3 Temporal Ctx	5.4	Control 4 Occipital Ctx	8.6
AD 4 Temporal Ctx	65.5	Control (Path) 1 Occipital Ctx	43.2
AD 5 Inf Temporal Ctx	44.4	Control (Path) 2 Occipital Ctx	6.6
AD 5 SupTemporal Ctx	38.4	Control (Path) 3 Occipital Ctx	35.4
AD 6 Inf Temporal Ctx	55.5	Control (Path) 4 Occipital Ctx	8.0
AD 6 Sup Temporal Ctx	64.2	Control 1 Parietal Ctx	10.8
Control 1 Temporal Ctx	12.9	Control 2 Parietal Ctx	39.8
Control 2 Temporal Ctx	27.4	Control 3 Parietal Ctx	12.4
Control 3 Temporal Ctx	18.6	Control (Path) 1 Parietal Ctx	35.1
Control 4 Temporal Ctx	10.9	Control (Path) 2 Parietal Ctx	14.8
Control (Path) 1 Temporal Ctx	41.2	Control (Path) 3 Parietal Ctx	43.5
Control (Path) 2 Temporal Ctx	28.5	Control (Path) 4 Parietal Ctx	20.6

Table MC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag6327, Run 259139893	issue Name	Rel. Exp.(%) Ag6327, Run 259139893
Adipose	16.7	Renal ca. TK-10	100.0

Melanoma* Hs688(A).T	12.1	Bladder	14.1
Melanoma* Hs688(B).T	15.0	Gastric ca. (liver met.) NCI-N87	29.9
Melanoma* M14	43.2	Gastric ca. KATO III	45.4
Melanoma* LOXIMVI	22.4	Colon ca. SW-948	2.9
Melanoma* SK-MEL-5	17.2	Colon ca. SW480	19.9
Squamous cell carcinoma SCC-4	18.4	Colon ca.* (SW480 met) SW620	16.0
Testis Pool	11.0	Colon ca. HT29	5.6
Prostate ca.* (bone met) PC-3	21.2	Colon ca. HCT-116	19.6
Prostate Pool	6.6	Colon ca. CaCo-2	11.1
Placenta	2.5	Colon cancer tissue	9.4
Uterus Pool	5.3	Colon ca. SW1116	3.2
Ovarian ca. OVCAR-3	8.8	Colon ca. Colo-205	3.2
Ovarian ca. SK-OV-3	25.3	Colon ca. SW-48	5.0
Ovarian ca. OVCAR-4	36.3	Colon Pool	8.8
Ovarian ca. OVCAR-5	23.2	Small Intestine Pool	7.1
Ovarian ca. IGROV-1	14.7	Stomach Pool	7.7
Ovarian ca. OVCAR-8	0.6	Bone Marrow Pool	3.6
Ovary	12.7	Fetal Heart	8.8
Breast ca. MCF-7	20.9	Heart Pool	3.8
Breast ca. MDA-MB-231	16.4	Lymph Node Pool	8.4
Breast ca. BT 549	40.1	Fetal Skeletal Muscle	4.9
Breast ca. T47D	6.8	Skeletal Muscle Pool	18.7
Breast ca. MDA-N	5.8	Spleen Pool	7.0
Breast Pool	11.3	Thymus Pool	8.2
Trachea	8.8	CNS cancer (glio/astro) U87-MG	7.5
Lung	2.5	CNS cancer (glio/astro) U-118-MG	17.7
Fetal Lung	69.7	CNS cancer (neuro;met) SK-N-AS	36.6
Lung ca. NCI-N417	3.7	CNS cancer (astro) SF-539	5.8
Lung ca. LX-1	15.5	CNS cancer (astro) SNB-75	27.5
Lung ca. NCI-H146	48.3	CNS cancer (glio) SNB-19	14.8
Lung ca. SHP-77	24.1	CNS cancer (glio) SF-295	33.2
Lung ca. A549	13.8	Brain (Amygdala) Pool	13.8
Lung ca. NCI-H526	3.8	Brain (cerebellum)	50.0
Lung ca. NCI-H23	20.0	Brain (fetal)	14.8
Lung ca. NCI-H460	18.9	Brain (Hippocampus) Pool	14.8
Lung ca. HOP-62	26.8	Cerebral Cortex Pool	21.5
Lung ca. NCI-H522	28.9	Brain (Substantia nigra) Pool	15.7
Liver	0.6	Brain (Thalamus) Pool	26.6
Fetal Liver	25.0	Brain (whole)	25.3
Liver ca. HepG2	8.4	Spinal Cord Pool	13.4
Kidney Pool	5.0	Adrenal Gland	6.5
Fetal Kidney	27.7	Pituitary gland Pool	1.6
Renal ca. 786-0	54.7	Salivary Gland	2.1
Renal ca. A498	19.1	Thyroid (female)	13.1

Renal ca. ACHN	20.3	Pancreatic ca. CAPAN2	19.2
Renal ca. UO-31	35.4	Pancreas Pool	11.6

Table MD. Panel 4.1D

Tissue Name	Rel. Exp.() Ag6327, Run 259181431	Tissue Name	Rel. Exp.(%) Ag6327, Run 259181431
Secondary Th1 act	11.7	HUVEC IL-1beta	9.8
Secondary Th2 act	20.6	HUVEC IFN gamma	8.7
Secondary Tr1 act	3.1	HUVEC TNF alpha + IFN gamma	1.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.2
Secondary Th2 rest	0.2	HUVEC IL-11	4.1
Secondary Tr1 rest	0.0	Lung Microvascular EC none	9.5
Primary Th1 act	0.1	Lung Microvascular EC TNFalpha + IL-1beta	2.9
Primary Th2 act	14.6	Microvascular Dermal EC none	0.8
Primary Tr1 act	11.0	Microvascular Dermal EC TNFalpha + IL-1beta	3.5
Primary Th1 rest	0.3	Bronchial epithelium TNFalpha + IL1beta	4.8
Primary Th2 rest	0.5	Small airway epithelium none	2.1
Primary Tr1 rest	0.3	Small airway epithelium TNFalpha + IL-1beta	8.3
CD45RA CD4 lymphocyte act	16.4	Coronary artery SMC rest	6.0
CD45RO CD4 lymphocyte act	19.8	Coronary artery SMC TNFalpha + IL-1beta	10.0
CD8 lymphocyte act	1.9	Astrocytes rest	1.3
Secondary CD8 lymphocyte rest	8.0	Astrocytes TNFalpha + IL-1beta	1.4
Secondary CD8 lymphocyte act	0.5	KU-812 (Basophil) rest	31.9
CD4 lymphocyte none	0.4	KU-812 (Basophil) PMA/ionomycin	40.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.6	CCD1106 (Keratinocytes) none	5.9
LAK cells rest	10.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	14.9
LAK cells IL-2	2.1	Liver cirrhosis	4.5
LAK cells IL-2+IL-12	0.4	NCI-H292 none	8.2
LAK cells IL-2+IFN gamma	3.9	NCI-H292 IL-4	7.1
LAK cells IL-2+ IL-18	1.4	NCI-H292 IL-9	11.0
LAK cells PMA/ionomycin	**100.0**	NCI-H292 IL-13	13.1
NK Cells IL-2 rest	16.0	NCI-H292 IFN gamma	5.5
Two Way MLR 3 day	3.6	HPAEC none	2.0
Two Way MLR 5 day	0.4	HPAEC TNF alpha + IL-1 beta	10.6
Two Way MLR 7 day	2.0	Lung fibroblast none	7.7

PBMC rest	0.6	Lung fibroblast TNF alpha + IL-1 beta	9.3
PBMC PWM	3.7	Lung fibroblast IL-4	4.8
PBMC PHA-L	2.1	Lung fibroblast IL-9	8.3
Ramos (B cell) none	1.3	Lung fibroblast IL-13	0.7
Ramos (B cell) ionomycin	7.7	Lung fibroblast IFN gamma	9.0
B lymphocytes PWM	5.9	Dermal fibroblast CCD1070 rest	9.5
B lymphocytes CD40L and IL-4	9.6	Dermal fibroblast CCD1070 TNF alpha	19.2
EOL-1 dbcAMP	10.4	Dermal fibroblast CCD1070 IL-1 beta	14.8
EOL-1 dbcAMP PMA/ionomycin	1.2	Dermal fibroblast IFN gamma	4.1
Dendritic cells none	15.2	Dermal fibroblast IL-4	9.4
Dendritic cells LPS	4.5	Dermal Fibroblasts rest	3.7
Dendritic cells anti-CD40	3.1	Neutrophils TNFa+LPS	10.5
Monocytes rest	0.5	Neutrophils rest	3.3
Monocytes LPS	36.9	Colon	0.6
Macrophages rest	5.1	Lung	0.7
Macrophages LPS	3.8	Thymus	0.9
HUVEC none	5.6	Kidney	5.1
HUVEC starved	6.5		

Table ME. Panel CNS 1.1

Tissue Name	Rel. Exp.(%) Ag6327, Run 259229795	Tissue Name	Rel. Exp.(%) Ag6327, Run 259229795
Cing Gyr Depression2	11.6	BA17 PSP2	6.3
Cing Gyr Depression	10.0	BA17 PSP	25.0
Cing Gyr PSP2	5.2	BA17 Huntington's2	8.7
Cing Gyr PSP	19.3	BA17 Huntington's	12.2
Cing Gyr Huntington's2	13.9	BA17 Parkinson's2	11.0
Cing Gyr Huntington's	40.9	BA17 Parkinson's	43.2
Cing Gyr Parkinson's2	19.9	BA17 Alzheimer's2	1.7
Cing Gyr Parkinson's	64.2	BA17 Control2	13.1
Cing Gyr Alzheimer's2	4.5	BA17 Control	24.0
Cing Gyr Alzheimer's	26.1	BA9 Depression2	11.9
Cing Gyr Control2	9.3	BA9 Depression	6.0
Cing Gyr Control	47.6	BA9 PSP2	1.6
Temp Pole Depression2	9.9	BA9 PSP	10.7
Temp Pole PSP2	3.5	BA9 Huntington's2	17.6
Temp Pole PSP	5.3	BA9 Huntington's	36.6
Temp Pole Huntington's	13.2	BA9 Parkinson's2	28.7

Temp Pole Parkinson's2	11.0	BA9 Parkinson's	54.0
Temp Pole Parkinson's	41.5	BA9 Alzheimer's2	5.1
Temp Pole Alzheimer's2	5.8	BA9 Alzheimer's	5.6
Temp Pole Alzheimer's	3.2	BA9 Control2	29.3
Temp Pole Control2	29.7	BA9 Control	8.3
Temp Pole Control	8.5	BA7 Depression	11.0
Glob Palladus Depression	7.1	BA7 PSP2	20.4
Glob Palladus PSP2	5.4	BA7 PSP	47.0
Glob Palladus PSP	6.5	BA7 Huntington's2	36.9
Glob Palladus Parkinson's2	10.7	BA7 Huntington's	36.1
Glob Palladus Parkinson's	100.0	BA7 Parkinson's2	11.3
Glob Palladus Alzheimer's2	8.3	BA7 Parkinson's	25.3
Glob Palladus Alzheimer's	17.0	BA7 Alzheimer's2	4.5
Glob Palladus Control2	8.1	BA7 Control2	9.0
Glob Palladus Control	17.7	BA7 Control	16.0
Sub Nigra Depression2	5.5	BA4 Depression2	4.7
Sub Nigra Depression	7.1	BA4 Depression	15.2
Sub Nigra PSP2	2.6	BA4 PSP2	12.9
Sub Nigra Huntington's2	18.8	BA4 PSP	9.5
Sub Nigra Huntington's	39.8	BA4 Huntington's2	5.2
Sub Nigra Parkinson's2	41.5	BA4 Huntington's	19.5
Sub Nigra Alzheimer's2	6.4	BA4 Parkinson's2	14.5
Sub Nigra Control2	19.6	BA4 Parkinson's	58.6
Sub Nigra Control	43.5	BA4 Alzheimer's2	2.5
BA17 Depression2	10.2	BA4 Control2	10.8
BA17 Depression	11.2	BA4 Control	12.7

CNS_neurodegeneration_v1.0 Summary: Ag6327 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. This gene is found to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

General_screening_panel_v1.5 Summary: Ag6327 Highest expression of this gene is detected in a renal cancer TK-10 cell line (CT=28.9). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung,

liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Interestingly, this gene is expressed at much higher levels in fetal (CTs=29-31) when compared to adult lung and liver (CT=34-36). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung and liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung and liver related diseases.

Panel 4.1D Summary: Ag6327 Highest expression of this gene is detected in PMA/ionomycin treated LAK cells (CT=27.8). Lower levels of expression of this gene is also seen in resting and cytokine treated LAK cells. These cells are involved in tumor immunology and cell clearance of virally and bacterial infected cells as well as tumors. Therefore, modulation of the function of the protein encoded by this gene through the application of a small molecule drug or antibody may alter the functions of these cells and lead to improvement of symptoms associated with these conditions.

Moderate to low levels of this gene is also seen in activated polarized, memory and naive T cells, IL-2 treated NK cells, two way MLR, activated PBMC cells, Ramos B cells, B lymphocytes, eosinophils, dendritic cells, activated monocytes, macrophages, endothelial cells, bronchial and small airway epithelium, mucoepidermoid cells, astrocytes, basophils, keratinocytes, lung and dermal fibroblasts, neutrophils and normal tissues represented by lung, thymus and kidney. Therefore, therapeutic modulation of this gene product may

ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Panel CNS_1.1 Summary: This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. Please see Panel 1.5 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and

protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome Research. 10 (8) 1249-1265, 2000).

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

NOV2 CG126119-02 SNP data:

CG126119-02 has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG126119-02 variant differ as shown in Table DJ.

Table DJ. cSNP and Coding Variants for CG126119-02.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380802	99	T	C	30	Ala	Ala
13380813	140	A	C	44	Asp	Ala

NOV3 CG137623-01 SNP data

CG137623-01 has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG137623-01 variant differ as shown in Table DB.

Table DB. cSNP and Coding Variants for CG137623-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380806	407	T	C	64	Leu	Pro
13380805	588	T	C	124	Thr	Thr

NOV5 CG143198-01 SNP data:

CG143198-01 has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG143198-01 variant differ as shown in Table DC.

Table DC. cSNP and Coding Variants for CG143198-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
cg111.5225	209	G	T	70	Gly	Val
cg111.5224	212	G	T	71	Arg	Ile

NOV6 CG144756-01 SNP data:

CG144756-01 has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG144756-01 variant differ as shown in Table DD.

Table DD. cSNP and Coding Variants for CG144756-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380801	480	C	T	153	Leu	Leu

NOV7 CG145473-01 SNP data:

CG145473-01 has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG145473-01 variant differ as shown in Table DE.

Table DE. cSNP and Coding Variants for CG145473-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
c100.13	459	T	C	131	Met	Thr
13380809	781	C	T	238	Ser	Ser

NOV9 CG146452-01 SNP data:

CG146452-01 has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG146452-01 variant differ as shown in Table DF.

Table DF. cSNP and Coding Variants for CG146452-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377204	52	A	G	0		
13377208	334	T	A	93	Ser	Thr
13377209	385	C	T	110	Leu	Leu

NOV11 CG147048-01 SNP data:

CG147048-01 has 6 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG147048-01 variant differ as shown in Table DG.

Table DG. cSNP and Coding Variants for CG147048-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380827	40	A	G	14	Thr	Ala
13380828	45	T	C	15	Cys	Cys
13380829	379	C	T	127	His	Tyr

13380830	406	C	G	136	His	Asp
13380831	484	G	A	162	Gly	Arg
13380832	1151	T	C	384	Phe	Ser

NOV12 CG147246-01 SNP data:

CG147246-01 has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG147246-01 variant differ as shown in Table DH.

Table DH. cSNP and Coding Variants for CG147246-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380833	435	C	T	145	Asp	Asp

NOV16 CG150951-01 SNP data

CG150951-01 has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG150951-01 variant differ as shown in Table DA.

Table DA. cSNP and Coding Variants for CG150951-01.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13381070	454	A	G	98	Glu	Gly

NOV20 CG59323-02 SNP data:

CG59323-02 has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: and , respectively. The nucleotide sequence of the CG59323-02 variant differ as shown in Table DI.

Table DI. cSNP and Coding Variants for CG59323-02.

Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380818	2843	A	G	758	Asp	Gly
13380819	2996	A	G	809	Glu	Gly

Example E: Protein Interactions in the DAPK3 (zip kinase) Signaling Pathway

Novel associations between Death Associated Protein Kinase 3 ("DAPK3") proteins and DAPK3 interacting polypeptides ("DAPK-IP"), and the nucleic acids that encode them, are described, as are various diseases or pathologies associated with DAPK3 and DAPK-IP protein complexes ("DAPK:DAPK-IP"). The DAPK-IP proteins, polypeptides and their cognate nucleic acids were identified by Curagen Corporation in certain cases. The DAPK-IP and any variants thereof, are suitable as diagnostic markers, targets for an antibody therapeutic and targets for a small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed DAPK:DAPK-IP protein complexes in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

DAPK3 is a serine/threonine kinase, designated ZIP kinase, that mediates apoptosis. ZIP kinase contains a leucine zipper structure at its C-terminus, and a kinase domain at its N-terminus. ZIP kinase physically binds to ATF4, a member of the activating transcription factor/cyclic AMP-responsive element-binding protein (ATF / CREB) family, through interaction between their leucine zippers. The leucine zipper domain is necessary for the homodimerization of ZIP kinase as well as for the activation of kinase. An immunostaining study showed that ZIP kinase localizes in the nuclei. Overexpression of intact ZIP kinase (but not catalytically inactive kinase mutants) led to the morphological changes of apoptosis in NIH 3T3 cells, suggesting that the cell death-inducing activity of ZIP kinase depends on its intrinsic kinase activity. Interestingly, the catalytic domain of ZIP kinase is closely related to that of death-associated protein kinase (DAP kinase), which is a mediator of apoptosis induced by gamma interferon. Therefore, both ZIP and DAP kinases represent a novel kinase family, which mediates apoptosis through their catalytic activities. PMID: 9488481 The ZIPK gene was mapped to 19q13.3 by fluorescence *in situ* hybridization and by polymerase chain reaction-based analyses with both a human/rodent monochromosomal hybrid cell panel and a radiation hybrid mapping panel.

ATF4(CREB2) is a stress-inducible gene. The bZIP domain of ATF4 forms a heterodimer with the bZIP domain of C/EBP beta that binds the cAMP response element, but not CCAAT box DNA, with high affinity. The basic region of ATF4 has a higher alpha-helical propensity than that of C/EBP beta. The degree of ordering of the basic region and the fork and the dimerization properties of the leucine zipper combine to distinguish

the structurally similar bZIP domains of ATF4 and C/EBP beta with respect to DNA target sequence. PMID: 11018027

CCAAT/enhancer binding protein-delta ("CEBPD") is important in the transcriptional activation and regulation of genes involved in immune and inflammatory responses. It may also play an important role in the regulation of the several genes associated with activation and /or differentiation of macrophages.

Gadd153, also known as chop, encodes a member of the CCAAT/enhancer-binding protein (C/EBP) transcription factor family and is transcriptionally activated by cellular stress signals. GADD153 inhibits the DNA-binding activity of CEBPD by forming heterodimers that cannot bind DNA. Dysregulation of GADD153 is seen in a form of myxoid liposarcoma. Arsenite treatment of rat pheochromocytoma PC12 cells results in the biphasic induction of Gadd153 mRNA expression, controlled in part through binding of C/EBPbeta and two uncharacterized protein complexes to the C/EBP-ATF (activating transcription factor) composite site in the Gadd153 promoter.

Components of these additional complexes are two ATF/CREB (cAMP-responsive-element-binding protein) transcription factors having differential binding activities dependent upon the time of arsenite exposure. During arsenite treatment of PC12 cells, enhanced binding of ATF4 to the C/EBP-ATF site at 2 h was observed as Gadd153 mRNA levels increased, and enhanced binding of ATF3 complexes at 6 h was observed as Gadd153 expression declined. ATF4 activates, while ATF3 represses, Gadd153 promoter activity through the C/EBP-ATF site. ATF3 also repressed ATF4-mediated transactivation and arsenite-induced activation of the Gadd153 promoter. Results suggest that numerous members of the ATF/CREB family are involved in the cellular stress response, and that regulation of stress-induced biphasic Gadd153 expression in PC12 cells involves the ordered, sequential binding of multiple transcription factor complexes to the C/EBP-ATF composite site. PMID: 10085237

CHOP/gadd153 is a transcription factor induced by cellular stresses such as UV light, genotoxic agents, and protein misfolding in the endoplasmic reticulum. These stresses induce CHOP expression, and at the same time cause cellular apoptosis. CHOP can directly induce apoptosis. A GFP-tagged CHOP vector, ectopically overexpressed in several cell types (3T3 fibroblasts, keratinocytes, and HeLa cells), caused apoptosis as defined by morphology, DNA fragmentation, and FACS analysis. Apoptosis was quantified using a rapid fluorescence assay that measures the signal from cells collected in culture

supernatants. Simultaneous overexpression of CHOP and p38 significantly augmented apoptosis. However, although p38 kinase clearly modulated the activity of full-length CHOP, it was not absolutely required. Deletion mapping experiments showed that the bZIP region of CHOP stimulates apoptosis to nearly the same extent as wild-type CHOP. Thus, while the amino-terminal region of CHOP serves an important modulatory role (i.e., regulation by p38), the underlying apoptosis-inducing activity of CHOP resides within the bZIP region of the molecule. PMID: 11426938

Hepatitis delta virus (HDV) is a pathogenic human virus whose RNA genome and replication cycle resemble those of plant viroids. However, viroid genomes contain no open reading frames, whereas HDV RNA encodes a single protein, hepatitis delta antigen (HDAg), which is required for viral replication. A cellular gene whose product interacts with HDAg has been identified, and this interaction was found to affect viral genomic replication in intact cells. DNA sequence analysis revealed that this protein, termed delta-interacting protein A (DIPA), is a cellular homolog of HDAg. These observations demonstrated that a host gene product can modulate HDV replication and suggested that HDV evolved from a primitive viroid-like RNA through capture of a cellular transcript. PMID: 8810253

The CEBPD inducible protein has an immunoglobulin ("Ig") domain as determined by Pfam analysis, and is predicted to be cytoplasmic. Ig domains are important mediators of protein-protein interactions. The interaction of CEBPD with ATF4 could indicate a role in ATF4-dependent transcription in response to cellular stress or cell proliferation.

In pancreatic acinar cells, the HOX-like factor PDX1 acts as part of a trimeric complex with two TALE class homeodomain factors, PBX1b and MEIS2b. The complex binds to overlapping half-sites for PDX1 and PBX. The trimeric complex activates transcription in cells to a level about an order of magnitude greater than PDX1 alone. The N-terminal PDX1 activation domain is required for detectable transcriptional activity of the complex, even though PDX1 truncations bearing only the PDX1 C-terminal homeodomain, and pentapeptide motifs can still participate in forming the trimeric complex. The conserved N-terminal PBC-B domain of PBX, as well as its homeodomain, is required for both complex formation and transcriptional activity. Only the N-terminal region of MEIS2, including the conserved MEIS domains, is required for formation of a trimer on DNA and transcriptional activity: the MEIS homeodomain is dispensable. The activity of the pancreas-specific ELA1 enhancer requires the cooperation of the trimer-binding element

and a nearby element that binds the pancreatic transcription factor PTF1. The PDX1. PBX1b.MEIS2b complex cooperates with the PTF1 basic helix-loop-helix complex to activate an ELA1 mini-enhancer in HeLa cells. This cooperation requires all three homeoprotein subunits, including the PDX1 activation domain. PMID: 11279116

Tumor endothelial factor 1 (TEM1; ENDOSIALIN) is a Ia-type plasma membrane protein with extracellular EGF and sushi domains. TEM1 is expressed by tumor blood vessel endothelium. Human tumor endothelial marker 1 /endosialin (TEM1/endosialin) was recently identified as a novel tumor endothelial cell surface marker potentially involved in angiogenesis, although no specific function for this novel gene has been assigned so far. It was reported to be expressed in tumor endothelium but not in normal endothelium with the exception of perhaps the corpus luteum. The mouse Tem1/endosialin homolog was identified and its promoter region was characterized. Its expression pattern in murine and human tissues and murine cell lines in vitro has been extensively characterized. The single copy gene that was mapped to the chromosome 19, is intronless and encodes a 92 kDa protein that has 77.5% overall homology to the human protein. This gene is ubiquitously expressed in normal human and mouse somatic tissues and during development, and its expression at the mRNA level is density dependent and upregulated in serum starved cells. In vitro, its expression is limited to cells of embryonic, endothelial and preadipocyte origin suggesting that the wide distribution of its expression in vivo is due to the presence of vascular endothelial cells in all the tissues. The ubiquitous expression in vivo is in contrast to previously reported expression limited to corpus luteum, and highly angiogenic tissues such as tumors and wound tissue. PMID: 11489895

Discovery Method:

DAPK-IPs were identified using PathCalling™ Technology (CuraGen Corporation). The sequence were derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were sequenced. In silico prediction was based on sequences available in Curagen Corporation's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states

from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E. coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

Description of the Interaction:

Death Associated Protein Kinases are very important in the propagation of the apoptotic signal due to the binding of Interferon gamma (INFg) to its receptor. These kinases constitute a novel family of protein kinases whose catalytic activity plays a critical

role in cell survival. In these Yeast-2-Hybrid experiments, the aa403-453 region of DAPK3 was used to screen for interacting proteins. The C-terminus of DAPK3 contains the leucine zipper motif that is critical for protein-protein interactions.

Sites of DAPK:DAPK-IP protein-protein interactions were mapped. Through its phosphorylating activity and leucine zipper protein-protein interaction motif, the DAPK3 interacted with ATF4 (aa5-,14-,71-,77-,80-351). ATF4 (aa2-, 7-, 66-, 71-351) interacted with CEBPD (aa211-281). The DAPK3/ATF4 interaction could modulate the CEBPD/ATF4 interaction with protein and/or DNA targets or modulate CEBPD-interactions with other proteins and/or DNA binding sites to regulate expression of various target genes.

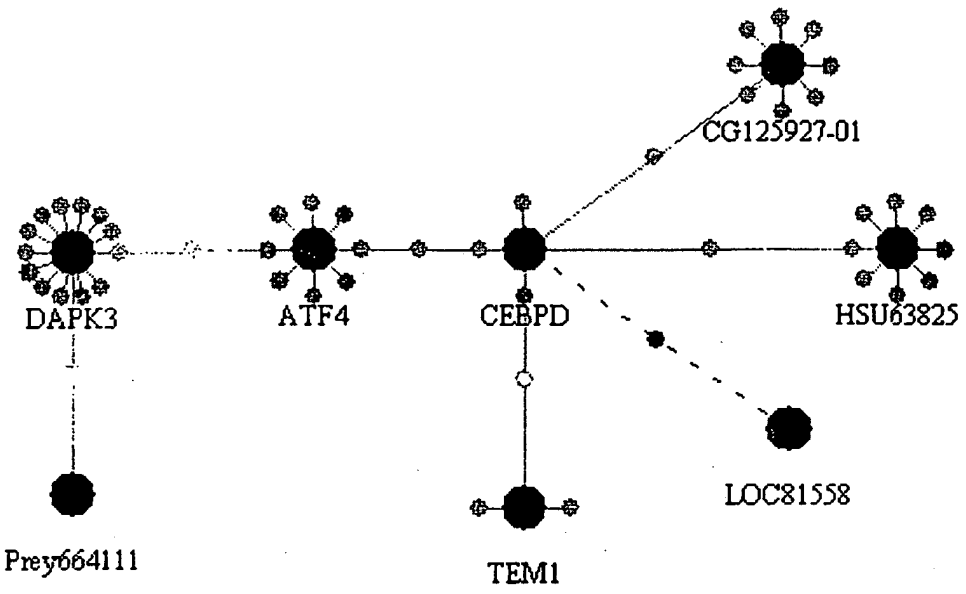
One target gene could be the CEBP-induced protein, whose function in cell survival is not entirely clear, but seems to indicate a possible feedback mechanism with CEBPD affecting its activity. The domains used in this interaction were; CEBPD (aa211-281), and CEBP-inducible protein (aa42-453). This is also observed in the interaction with GADD153, which is a transcription co-regulator that binds to CEBPD and prevents it from binding DNA on the promoters of target genes. The domains that gave the interaction are CEBPD (aa211-281), and GADD153 (aa40-, 43-, 51-169). Gadd153 expression is increased due to growth arrest and DNA-damage and promotes apoptosis.

TEM1 is a type Ia plasma membrane protein with extracellular EGF and sushi domains. These domains are important in transducing cell signaling events such as in growth, differentiation, and apoptosis. The C-terminus of TEM1 was used in the Yeast-2-Hybrid screen. This domain likely contains motifs important for transducing cellular signals. TEM1 is expressed by tumor blood vessel endothelium and could be receptor, the endogenous ligand is as of yet unknown. The interaction of CEBPD with TEM1 could indicate a signaling step in binding of TEM1 to its ligand. The domains that gave this interaction were; CEBPD (aa211-281), and TEM1 (aa686-757). CEBPD is particularly interesting because of its importance in inflammatory signaling cascades. The interaction of DAPK3 with the homeobox protein, MEIS2, may function to alter its activity and modulate the expression of genes during development or house-keeping type genes in the adult. Additionally, the interaction of the DIPA protein, which has been shown to interact with viral phosphoprotein Hepatitis Delta Antigen (HDA), with CEBPD could indicate an interaction necessary for the disruption of the INFg apoptotic signal allowing

for increased viral replication. The domains that gave the interaction were; CEBPD (aa211-281), and DIPA (aa31-, 52-202).

The identified DAPK3 protein interactions and their corresponding biological pathways are shown in Table F1.

Table F1: Diagram of Protein Interactions and/or Biological Pathway:



The nucleotide and polypeptide sequences of the identified DAPK-IPs are provided in Table F2.

Table F2. Nucleic Acid and Amino Acid Sequence Analysis of Interacting Proteins

	SEQ ID NO: 106	2565 bp	
TEM1 DNA Sequence	TCGCGATGCTGCTGCGCCTGTTGCTGGCCTGGGCGGCCGCAGGGCCCACACTGGG CCAGGACCCCTGGGCTGCTGAGCCCCGTGCCGCCTGCGGCCCCAGCAGCTGCTAC GCTCTCTTCCCACGGCGCCGCACCTTCCTGGAGGCCTGGCGGGCCTGCCGCGAGC TGGGGGGCGACCTGGCCACTCCTCGGACCCCCGAGGAGGCCCAGCGTGTGGACAG CCTGGTGGGTGCGGGCCCAGCCAGCCGGCTGCTGTGGATCGGGCTGCAGCGGCAG GCCCGGCAATGCCAGCTGCAGCGCCCACTGCGCGGCTTCACGTGGACCACAGGGG ACCAGGACACGGCTTTCACCAACTGGGCCCAGCCAGCCTCTGGAGGCCCCTGCCC GGCCCAGCGCTGTGTGGCCCTGGAGGCAAGTGGCGAGCACCGCTGGCTGGAGGGC TCGTGCACGCTGGCTGTCGACGGCTACCTGTGCCAGTTTGGCTTCGAGGGCGCCT GCCCGGCGCTGCAAGATGAGGCGGGCCAGGCCGGCCCAGCCGTGTATACCACGCC CTTCCACCTGGTCTCCACAGAGTTTGAGTGGCTGCCCTTCGGCTCTGTGGCCGCT GTGCAGTGCCAGGCTGGCAGGGGAGCCTCTCTGCTCTGCGTGAAGCAGCCTGAGG GAGGTGTGGGCTGGTCACGGGCTGGGCCCCTGTGCCTGGGGACTGGCTGCAGCCC TGACAACGGGGGCTGCGAACACGAATGTGTGGAGGAGGTGGATGGTCACGTGTCC		

	TGCCGCTGCACTGAGGGCTTCCGGCTGGCAGCAGACGGGCGCAGTTGCGAGGACC CCTGTGCCCAGGCTCCGTGCGAGCAGCAGTGTGAGCCCGGTGGGCCACAAGGCTA CAGCTGCCACTGTCGCCTGGGTTTCCGGCCAGCGGAGGATGATCCGCACCGCTGT GTGGACACAGATGAGTGCCAGATTGCCGGTGTGTGCCAGCAGATGTGTGTCAACT ACGTTGGTGGCTTCGAGTGTTATTGTAGCGAGGGACATGAGCTGGAGGCTGATGG CATCAGCTGCAGCCCTGCAGGGGCCATGGGTGCCCAGGCTTCCCAGGACCTCGGA GATGAGTTGCTGGATGACGGGGAGGATGAGGAAGATGAAGACGAGGCCTGGAAGG CCTTCAACGGTGGCTGGACGGAGATGCCTGGGATCCTGTGGATGGAGCCTACGCA GCCGCCTGACTTTGCCCTGGCCTATAGACCGAGCTTCCCAGAGGACAGAGAGCCA CAGATACCCTACCCGGAGCCCACCTGGCCACCCCCGCTCAGTGCCCCCAGGGTCC CCTACCACTCCTCAGTGCTCTCCGTCACCCGGCCTGTGGTGGTCTCTGCCACGCA TCCCACACTGCCTTCTGCCCACCAGCCTCCTGTGATCCCTGCCACACACCCAGCT TTGTCCCGTGACCACCAGATCCCCGTGATCGCAGCCAACTATCCAGATCTGCCTT CTGCCTACCAACCCGGTATTCTCTCTGTCTCTCATTCAGCACAGCCTCCTGCCCA CCAGCCCCCTATGATCTCAACCAAATATCCGGAGCTCTTCCCTGCCCACCAGTCC CCCATGTTTCCAGACACCCGGGTCGCTGGCACCCAGACCACCACTCATTTGCCTG GAATCCCACCTAACCATGCCCCTCTGGTCACCACCCTCGGTGCCCAGCTACCCCC TCAAGCCCCAGATGCCCTTGTCCTCAGAACCCAGGCCACCCAGCTTCCCATTATC CCAACTGCCCAGCCCTCTCTGACCACCACCTCCAGGTCCCCTGTGTCTCCTGCCC ATCAAATCTCTGTGCCTGCTGCCACCCAGCCCGCAGCCCTCCCCACCCTCCTGCC CTCTCAGAGCCCCACTAACCAGACCTCACCCATCAGCCCTACACATCCCCATTCC AAAGCCCCCCAAATCCCAAGGGAAGATGGCCCCAGTCCCAAGTTGGCCCTGTGGC TGCCCTCACCAGCTCCCACAGCAGCCCCAACAGCCCTGGGGGAGGCTGGTCTTGC CGAGCACAGCCAGAGGGATGACCGGTGGCTGCTGGTGGCACTCCTGGTGCCAACG TGTGTCTTTTTGGTGGTCCTGCTTGCACTGGGCATCGTGTACTGCACCCGCTGTG GCCCCCATGCACCCAACAAGCGCATCACTGACTGCTATCGCTGGGTCATCCATGC TGGGAGCAAGAGCCCAACAGAACCCATGCCCCCCAGGGGCAGCCTCACAGGGGTG CAGACCTGCAGAACCAGCGTGTGATGGGGTGCAGACCCCCCTCATGGAGTATGGG GCGCTGGACACATGGCCGGGGCTGCACCAGGGACCCATGGGGGCTGCCCAGCTGG ACAGATGGCTTCCTGCTCCCCAGGCCCAGCCAGGGTCCTCTCTCAACCACTAGAC TTGGCTCTCAGGAACTCTGCTTCCTGGCCCAGCGCTCGTGACCAAGGATACACCA AAGCCCTTAAGACCTCAGGGGGCGGGTGCTGGGGTCTTCTCCAATAAATGGGGTG TCAACCTTAAAAAAAAAAAAAAAAAAAAAAAAAAA
	SEQ ID NO: 107 | 757 aa
TEM1 Protein Sequence	MLLRLLLAWAAAGPTLGQDPWAAEPRAACGPSSCYALFPRRRTFLEAWRACRELG GDLATPRTPEEAQRVDSLVGAGPASRLLWIGLQRQARQCQLQRPLRGFTWTTGDQ DTAFTNWAQPASGGPCPAQRCVALEASGEHRWLEGSCTLAVDGYLCQFGFEGACP ALQDEAGQAGPAVYTTPFHLVSTEFEWLPFGSVAAVQCQAGRGASLLCVKQPEGG VGWSRAGPLCLGTGCSPDNGGCEHECVEEVDGHVSCRCTEGFRLAADGRSCEDPC AQAPCEQQCEPGGPQGYSCHCRLGFRPAEDDPHRCVDTDECQIAGVCQQMCVNYV GGFECYCSEGHELEADGISCSPAGAMGAQASQDLGDELLDDGEDEEDEDEAWKAF NGGWTEMPGILWMEPTQPPDFALAYRPSFPEDREPQIPYPEPTWPPPLSAPRVPY HSSVLSVTRPVVVSATHPTLPSAHQPPVIPATHPALSRDHQIPVIAANYPDLPSA YQPGILSVSHSAQPPAHQPPMISTKYPELFPAHQSPMFPDTRVAGTQTTTHLPGI PPNHAPLVTTLGAQLPPQAPDALVLRTQATQLPIIPTAQPSLTTTSRSPVSPAHQ ISVPAATQPAALPTLLPSQSPTNQTSPISPTHPHSKAPQIPREDGPSPKLALWLP SPAPTAAPTALGEAGLAEHSQRDDRWLLVALLVPTCVFLVVLLALGIVYCTRCGP HAPNKRITDCYRWVIHAGSKSPTEPMPPRGSLTGVQTCRTSV
	SEQ ID NO: 108 | 2330 bp
CEBPD CG123869-01 DNA Sequence	GNCCCTCGACGGCNTGCAGCCGGGAGAGCCATGGCGGGGGCCGCAGCGGGCGGCA GAGGCGGAGGTGCCTGGGGGCCGGGGCGCGGAGGGGCCGGGGGGCTCCGGCGGGG CTGCTCTCCCCCAGCCCCCGCCGGCTCCCCCCGGGCTGGGCTGCAGCCGCTCAGG GCCACGATCCCCTTCCAGCTGCAGCAGCCGCACCAGCGCCGGGACGGGGGTGGCC GTGCAGCCAGCGTCCCATGCTCGGTGGCCCCAGAAAAGTCAGTGTGTAGGCCTCA

	GCCACTTCAGGTCCGGCGTACATTCTCCCTGGACACCATCCTCAGCTCCTACCTT CTGGGCCAGTGGCCACGAGATGCTGATGGGGCCTTCACCTGCTGCACCAATGACA AGGCCACCCAGACGCCCCTGTCCTGGCAAGAGCTAGAAGGTGAGCGTGCCAGTTC CTGTGCACACAAGCGCTCAGCATCCTGGGGCAGCACAGACCACCGAAAAGAGATT TCCAAGTTGAAGCAACAACTGCAGAGGACGAAGCTGAGCCGCAGTGGGAAAGAGA AGGAGCGAGGTTCACCACTCCTAGGGGACCACGCAGTGCGGGGAGCACTGAGGGC GTCCCCTCCCAGCTTCCCCTCAGGGTCCCCTGTCTTGCGACTCAGCCCCTGCCTG CACAGGAGCCTGGAAGGGCTCAACCAAGAGCTGGAGGAGGTATTTGTGAAGGAGC AGGGAGAAGAGGAGCTGCTGAGGATCCTTGATATCCCTGATGGGCACCGGGCCCC AGCTCCTCCCCAGAGTGGCAGCTGTGATCATCCCCTCCTCCTCCTGGAGCCTGGC AACCTTGCCAGCTCTCCTTCCATGTCCTTGGCATCTCCCCAGCCTTGTGGCCTGG CCAGTCATGAGGAACATCGGGGTGCCGCCGAGGAGCTGGCATCCACCCCCAACGA CAAAGCCTCCTCTCCAGGACACCCAGCCTTTCTTGAAGATGGCAGCCCATCTCCA GTCCTTGCCTTTGCTGCCTCCCCTCGACCTAATCATAGCTACATCTTCAAACGGG AGCCCCCAGAAGGCTGTGAGAAAGTGCGTGTGTTTGAAGAAGCCACGTCTCCAGG TCCTGACCTGGCCTTCCTGACTTCCTGTCCTGACAAGAACAAAGTCCATTTCAAC CCGACTGGCTCAGCCTTCTGCCCCGTCAACCTGATGAAGCCCCTCTTCCCCGGCA TGGGCTTCATCTTCCGTAACTGCCCCTCAAACCCGGGATCTCCCCTTCCCCCGGC CAGCCCCAGGCCACCACCTCGGAAGGATCCGGAAGCCTCCAAGGCCTCCCCACTG CCATTCGAGCCATGGCAGCGCACCCCACCATCAGAAGAGCCTGTGCTTTTCCAGA GCTCCCTGATGGTCTGAGGGTCCCACCCCTGCCCCACTTTACCATAGAGACCAGT GCCTTGGTGGCAGGTCCCTCCCCAGGTCCCCTGAGATGGGGTATGGAGGGGCCCT TCCCTCTCGGCCTTCGAGCACTTTCTTTCACTTACTGTGTCAAAGCCCTGGGTCC TCTTTTTGATGGGCACCGGCCCCTCTGAACGTGATGGGACCTGCCTTCTCCACTA GTAGCTGGGCAGCTCACAATTCACACCTGTGTACCTGCCACATCCCTCACTTGGT GGAAAACACCCAGAAGGTCTTGAGTCCCCCACCCCTGGGTGTCAGTCCAAATGAC TGTATAGGAGGCCCTTATTTTTGTCACAGAGCAAGCTGGCCATGAACGAAGGAGA GAAGACGCCACAGATTTCCTTCCCTCTCCTCCAGGAGACCATAAGATAGATCCCC CATCCTCTCAGCCCTATTCCCATGCCTCCCTCTCATTGGAGGAGCTGACCAAAGC AGCCCTAACGGGCCATAACACTTGACCAATTCAGCTGCTGGCAGAGGGAGGAAAC AAGTGTTTTCCCAAGTGGCATTTTCATCTCGCTTTCACCCTGACTAAAGATTGTC TTAAGTAGCAGCCCAGCCCGCCCAGCCCCAGGTGGGTAGTGGGGAGGAGAGCTGG CATTCCTCCAGGTGGCAAATGGCGACTCTATACTCTCCGCCCGCCCCAGGGCTGG ATGGATTAGAAAAATCCCTATTTTTCTTGTATCGATGTAGAGACTCTATTTTCTC CCAAAGACACTATTTTTGCAGCTGTTTGAAGTTTGTATATTTTCCGTACTGCAGA GCTTACACAAAATTGAAGAATGTTAATGTTCGAGTTTTCTTATCTTGTGTTTAGA GGTTGTTTTTTGCAGATCTTGGTGTTAATAGACCAAATAAATAAATAAATATTCC CAGCAAAAAAAAAAGTCGAC
	SEQ ID NO: 109 | 453 aa
CEBPD CG123869-01 Protein Sequence	MAGAAAGGRGGGAWGPGRGGAGGLRRGCSPPAPAGSPRAGLQPLRATIPFQLQQP HQRRDGGGRAASVPCSVAPEKSVCRPQPLQVRRTFSLDTILSSYLLGQWPRDADG AFTCCTNDKATQTPLSWQELEGERASSCAHKRSASWGSTDHRKEISKLKQQLQRT KLSRSGKEKERGSPLLGDHAVRGALRASPPSFPSGSPVLRLSPCLHRSLEGLNQE LEEVFVKEQGEEELLRILDIPDGHRAPAPPQSGSCDHPLLLLEPGNLASSPSMSL ASPQPCGLASHEEHRGAAEELASTPNDKASSPGHPAFLEDGSPSPVLAFAASPRP NHSYIFKREPPEGCEKVRVFEEATSPGPDLAFLTSCPDKNKVHFNPTGSAFCPVN LMKPLFPGMGFIFRNCPSNPGSPLPPASPRPPPRKDPEASKASPLPFEPWQRTPP SEEPVLFQSSLMV
	SEQ ID NO: 110 | 879 bp
DIPA CG129212-01 DNA Sequence	GGGCGATGCTCCAGAGGCCTGACCAGCCATGGAGGCCGAGGCAGGCGGCCTGGAG GAGCTGACGGACGAGGAGATGGCGGCGCTAGGCAAGGAAGAGCTAGTGCGGCGCC TGCGGCGGGAGGAGGCGACGCGCCTGGCGGCACTGGTGCAGCGCGGCCGCCTCAT GCAGGAGGTGAATCGGCAGCTGCAGGGCCACCTGGGCGAGATCCGCGAGCTCAAG CAGCTCAACCGGCGTCTGCAGGCAGAGAACCGTGAGCTGCGCGACCTCTGCTGCT

	TCCTGGACTCGGAGCGCCAGCGCGGGCGGCGCGCCGCACGCCAGTGGCAGCTCTT CGGGACCCAAGCATCCCGGGCCGTGCGCGAGGACCTGGGCGGCTGTTGGCAGAAG CTGGCCGAGCTGGAGGGCCGCCAGGAGGAGCTGCTGCGGGAGAACCTAGCGCTTA AGGAGCTCTGCCTGGCGCTGGGCGAAGAATGGGGCCCCCGCGGCGGCCCCAGCGG CGCCGGGGGATCAGGAGCCGGGCCAGCACCCGAGCTTGCCTTGCCCCCGTGCGGG CCCCGCGACCTAGGCGATGGAAGCTCCAGCACTGGCAGCGTGGGCAGTCCGGATC AGTTGCCCCTGGCCTGTTCCCCCGATGATTGAAGGCACTGCTTCCTCCACGCCGA CGCCCGCCCGGATTGCTCCCCGAGCCCCGGGACCGCTGTGGACCTCGGGACCTGG ACGCCGTCCTGGCTGCGCAGGAGGGGCCGCTGGCATGGACTAAGAAATCCTGACA CCAAGAAGGGCCCCTCGCTCTTGCTGGCAGGGCAGCAGGGGGACTGAAGGCTGGA GCGGAGGGACTTGCTGGGGGTTGGATTGGGGGTAATAAACCCGGACGGAAGCGG
	SEQ ID NO: 111 | 202 aa
DIPA CG129212-01 Protein Sequence	MEAEAGGLEELTDEEMAALGKEELVRRLRREEATRLAALVQRGRLMQEVNRQLQG HLGEIRELKQLNRRLQAENRELRDLCCFLDSERQRGRRAARQWQLFGTQASRAVR EDLGGCWQKLAELEGRQEELLRENLALKELCLALGEEWGPRGGPSGAGGSGAGPA PELALPPCGPRDLGDGSSSTGSVGSPDQLPLACSPDD
	SEQ ID NO: 112 | 965 bp
GADD153 CG125927-01 DNA Sequence	GGCACGAGGGAGAGAGAGAGACTTAAGTCTAAGGCACTGAGCGTATCATGTTAAA GATGAGCGGGTGGCAGCGACAGAGCCAAAATCAGAGCTGGAACCTGAGGAGAGAG TGTTCAAGAAGGAAGTGTATCTTCATACATCACCACACCTGAAAGCAGATGTGCT TTTCCAGACTGATCCAACTGCAGAGATGGCAGCTGAGTCATTGCCTTTCTCCTTT GGGACACTGTCCAGCTGGGAGCTGGAAGCCTGGTATGAGGACCTGCAAGAGGTCC TGTCTTCAGATGAAAATGGGGGTACCTATGTTTCACCTCCTGGAAATGAAGAGGA AGAATCAAAAATCTTCACCACTCTTGACCCTGCTTCTCTGGCTTGGCTGACTGAG GAGGAGCCAGAACCAGCAGAGGTCACAAGCACCTCCCAGAGCCCTCACTCTCCAG ATTCCAGTCAGAGCTCCCTGGCTCAGGAGGAAGAGGAGGAAGACCAAGGGAGAAC CAGGAAACGGAAACAGAGTGGTCATTCCCCAGCCCGGGCTGGAAAGCAGCGCATG AAGGAGAAAGAACAGGAGAATGAAAGGAAAGTGGCACAGCTAGCTGAAGAGAATG AACGGCTCAAGCAGGAAATCGAGCGCCTGACCAGGGAAGTAGAGGCGACTCGCCG AGCTCTGATTGACCGAATGGTGAATCTGCACCAAGCATGAACAATTGGGAGCATC AGTCCCCCACTTGGGCCACACTACCCACCTTTCCCAGAAGTGGCTACTGACTACC CTCTCACTAGTGCCAATGATGTGACCCTCAATCCCACATACGCAGGGGGAAGGCT TGGAGTAGACAAAAGGAAAGGTCTCAGCTTGTATATAGAGATTGTACATTTATTT ATTACTGTCCCTATCTATTAAAGTGACTTTCTATGAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAA
	SEQ ID NO: 113 | 169 aa
GADD153 CG125927-01 Protein Sequence	MSAALFSLDGPARGAPWPAEPAPFYEPGRAGKPGRGAEPGALGEPGAAAPAMYDD ESAIDFSAYIDSMAAVPTLELCHDELFADLFNSNHKAGGAGPLELLPGGPARPLG PGPAAPRLLKREPDWGDGDAPGSLLPAQVGPCAQTVVSLAAAGQPTPPTSPEPPR SSPRQTPAPGPAREKSAGKRGPDRGSPEYRQRRERNNIAVRKSRDKAKRRNQEMQ QKLVELSAENEKLHQRVEQLTRDLAGLRQFFKQLPSPPFLPAAGTADCR
	SEQ ID NO: 114 | 1594 bp
CEBPD DNA Sequence	CCCGGGGCGCCCCCGCGGTGCCGGAGTCGGGGCGGGGCGTGCACGTCAGCCGGGG CTAGAAAAGGCGGCGGGGCTGGGCCCAGCGAGGTGACAGCCTCGCTTGGACGCAG AGCCCGGCCCGACGCCGCCATGAGCGCCGCGCTCTTCAGCCTGGACGGCCCGGCG CGCGGCGCGCCCTGGCCTGCGGAGCCTGCGCCCTTCTACGAACCGGGCCGGGCGG GCAAGCCGGGCCGCGGGGCCGAGCCAGGGGCCCTAGGCGAGCCAGGCGCCGCCGC CCCCGCCATGTACGACGACGAGAGCGCCATCGACTTCAGCGCCTACATCGACTCC ATGGCCGCCGTGCCCACCCTGGAGCTGTGCCACGACGAGCTCTTCGCCGACCTCT TCAACAGCAATCACAAGGCGGGCGGCGCGGGGCCCCTGGAGCTTCTTCCCGGCGG CCCCGCGCGCCCCTTGGGCCCGGGCCCTGCCGCTCCCCGCCTGCTCAAGCGCGAG CCCGACTGGGGCGACGGCGACGCGCCCGGCTCGCTGTTGCCCGCGCAGGTGGGCC CGTGCGCACAGACCGTGGTGAGCTTGGCGGCCGCAGGGCAGCCCACCCCGCCCAC

	GTCGCCGGAGCCGCCGCGCAGCAGCCCCAGGCAGACCCCCGCGCCCGGCCCCGCC CGGGAGAAGAGCGCCGGCAAGAGGGGCCCGGACCGCGGCAGCCCCGAGTACCGGC AGCGGCGCGAGCGCAACAACATCGCCGTGCGCAAGAGCCGCGACAAGGCCAAGCG GCGCAACCAGGAGATGCAGCAGAAGTTGGTGGAGCTGTCGGCTGAGAACGAGAAG CTGCACCAGCGCGTGGAGCAGCTCACGCGGGACCTGGCCGGCCTCCGGCAGTTCT TCAAGCAGCTGCCCAGCCCGCCCTTCCTGCCGGCCGCCGGGACAGCAGACTGCCG GTAACGCGCGGCCGGGGCGGGAGAGACTCAGCAACGACCCATACCTCAGACCCGA CGGCCCGGAGCGGAGCGCGCCCTGCCCTGGCGCAGCCAGAGCCGCCGGGTGCCCG CTGCAGTTTCTTGGGACATAGGAGCGCAAAGAAGCTACAGCCTGGACTTACCACC ACTAAACTGCGAGAGAAGCTAAACGTGTTTATTTTCCCTTAAATTATTTTTGTAA TGGTAGCTTTTTCTACATCTTACTCCTGTTGATGCAGCTAAGGTACATTTGTAAA AAGAAAAAAAACCAGACTTTTCAGACAAACCCTTTGTATTGTAGATAAGAGGAAA AGACTGAGCATGCTCACTTTTTTATATTAATTTTTACAGTATTTGTAAGAATAAA GCAGCATTTGAAATCGCCCCTGCTTCCTATATTCGCAGTGACTCCCGCCCGCCCG CCGCCGCCGGTCGGAGGACCCGGCTCGGAAGGGCGTTCCGGACCGCAGCCAGCCA GCACCTAGGGAGCCCGGGCGCCAGGTGTGTGTGTGGGGGGGGCGGGGGGATGGGC GCAGCGGCGAGCTACTCAGGAGAGAGGGTCTGTCGCTTTTAAAACGCATTAAAGG CTCTCTCCTGGCCTTATTTAACTTGCCTAAGCTAGGTGGAGCACGGCTGAGCTC
	SEQ ID NO: 115 269 aa
CEBPD Protein Sequence	MSAALFSLDGPARGAPWPAEPAPFYEPGRAGKPGRGAEPGALGEPGAAAPAMYDD ESAIDFSAYIDSMAAVPTLELCHDELFADLFNSNHKAGGAGPLELLPGGPARPLG PGPAAPRLLKREPDWGDGDAPGSLLPAQVGPCAQTVVSLAAAGQPTPPTSPEPPR SSPRQTPAPGPAREKSAGKRGPDRGSPEYRQRRERNNIAVRKSRDKAKRRNQEMQ QKLVELSAENEKLHQRVEQLTRDLAGLRQFFKQLPSPPFLPAAGTADCR
	SEQ ID NO: 116 2015 bp
ATF4 DNA Sequence	GTTTTCTACTTTGCCCGCCCACAGATGTAGTTTTCTCTGCGCGTGTGCGTTTTCC CTCCTCCCCCGCCCTCAGGGTCCACGGCCACCATGGCGTATTAGGGGCAGCAGTG CCTGCGGCAGCATTGGCCTTTGCAGCGGCGGCAGCAGCACCAGGCTCTGCAGCGG CAACCCCCAGCGGCTTAAGCCATGGCGTGAGTACCGGGGCGGGTCGTCCAGCTGT GCTCCTGGGGCCGGCGCGGGTTTTGGATTGGTGGGGTGCGGCCTGGGGCCAGGGC GGTGCCGCCAAGGGGGAAGCGATTTAACGAGCGCCCGGGACGCGTGGTCTTTGCT TGGGTGTCCCCGAGACGCTCGCGTGCCTGGGATCGGGAAAGCGTAGTCGGGTGCC CGGACTGCTTCCCCAGGAGCCCTACAGCCCTCGGACCCCGAGCCCCGCAAGGTCC CAGGGGTCTTGGCTGTTGCCCCACGAAACGTGCAGGAACCAAGATGGCGGCGGCA GGGCGGCGGCGCGGGCGTGAGTCAAGGGCGGGCGGTGGGCGGGGCGCGGCCGCTG GCCGTATTTGGACGTGGGGACGGAGCGCTTTCCTCTTGGCGGCCGGTGGAAGAAT CCCCTGGTCTCCGTGAGCGTCCATTTTGTGGAACCTGAGTTGCAAGCAGGGAGGG GCAAATACAACTGCCCTGTTCCCGATTCTCTAGATGGCCGATCTAGAGAAGTCCC GCCTCATAAGTGGAAGGATGAAATTCTCAGAACAGCTAACCTCTAATGGGAGTTG GCTTCTGATTCTCATTCAGGCTTCTCACGGCATTCAGCAGCAGCGTTGCTGTAAC CGACAAAGACACCTTCGAATTAAGCACATTCCTCGATTCCAGCAAAGCACCGCAA CATGACCGAAATGAGCTTCCTGAGCAGCGAGGTGTTGGTGGGGGACTTGATGTCC CCCTTCGACCCGTCGGGTTTGGGGGCTGAAGAAAGCCTAGGTCTCTTAGATGATT ACCTGGAGGTGGCCAAGCACTTCAAACCTCATGGGTTCTCCAGCGACAAGGCTAA GGCGGGCTCCTCCGAATGGCTGGCTGTGGATGGGTTGGTCAGTCCCTCCAACAAC AGCAAGGAGGATGCCTTCTCCGGGACAGATTGGATGTTGGAGAAAATGGATTTGA AGGAGTTCGACTTGGATGCCCTGTTGGGTATAGATGACCTGGAAACCATGCCAGA TGACCTTCTGACCACGTTGGATGACACTTGTGATCTCTTTGCCCCCCTAGTCCAG GAGACTAATAAGCAGCCCCCCCAGACGGTGAACCCAATTGGCCATCTCCCAGAAA GTTTAACAAAACCCGACCAGGTTGCCCCCTTCACCTTCTTACAACCTCTTCCCCT TTCCCCAGGGGTCCTGTCCTCCACTCCAGATCATTCCTTTAGTTTAGAGCTGGGC AGTGAAGTGGATATCACTGAAGGAGATAGGAAGCCAGACTACACTGCTTACGTTG CCATGATCCCTCAGTGCATAAAGGAGGAAGACACCCCTTCAGATAATGATAGTGG CATCTGTATGAGCCCAGAGTCCTATCTGGGGTCTCCTCAGCACAGCCCCTCTACC

	AGGGGCTCTCCAAATAGGAGCCTCCCATCTCCAGGTGTTCTCTGTGGGTCTGCCC GTCCCAAACCTTACGATCCTCCTGGAGAGAAGATGGTAGCAGCAAAAGTAAAGGG TGAGAAACTGGATAAGAAGCTGAAAAAAATGGAGCAAAACAAGACAGCAGCCACT AGGTACCGCCAGAAGAAGAGGGCGGAGCAGGAGGCTCTTACTGGTGAGTGCAAAG AGCTGGAAAAGAAGAACGAGGCTCTAAAAGAGAGGGCGGATTCCCTGGCCAAGGA GATCCAGTACCTGAAAGATTTGATAGAAGAGGTCCGCAAGGCAAGGGGGAAGAAA AGGGTCCCCTAGTTGAGGATAGTCAGGAGCGTCAATGTGCTTGTACATAGAGTGC TGTAGCTGTGTGTTCCAATAAATTATTTTGTAGGG
	SEQ ID NO: 117 351 aa
ATF4 Protein Sequence	MTEMSFLSSEVLVGDLMSPFDPSGLGAEESLGLLDDYLEVAKHFKPHGFSSDKAK AGSSEWLAVDGLVSPSNNSKEDAFSGTDWMLEKMDLKEFDLDALLGIDDLETMPD DLLTTLDDTCDLFAPLVQETNKQPPQTVNPIGHLPESLTKPDQVAPFTFLQPLPL SPGVLSSTPDHSFSLELGSEVDITEGDRKPDYTAYVAMIPQCIKEEDTPSDNDSG ICMSPESYLGSPQHSPSTRGSPNRSLPSPGVLCGSARPKPYDPPGEKMVAAKVKG EKLDKKLKKMEQNKTAATRYRQKKRAEQEALTGECKELEKKNEALKERADSLAKE IQYLKDLIEEVRKARGKKRVP
	SEQ ID NO: 118 2132 bp
DAPK3 CG56543-01 DNA Sequence	GTTGCCATTAGGGGACTCCTGAGGTCCTATCTCCAGGCTGCGGTGACTGCACTTT CCCTGGAGTGGAAGCTGCTGGAAGGCGGACCGGCCGCCATGTCCACGTTCAGGCA GGAGGACGTGGAGGACCATTATGAGATGGGGGAGGAGCTGGGCAGCGGCCAGTTT GCGATCGTGCGGAAGTGCCGGCAGAAGGGCACGGGCAAGGAGTACGCAGCCAAGT TCATCAAGAAGCGCCGCCTGTCATCCAGCCGGCGTGGGGTGAGCCGGGAGGAGAT CGAGCGGGAGGTGAACATCCTGCGGGAGATCCGGCACCCCAACATCATCACCCTG CACGACATCTTCGAGAACAAGACGGACGTGGTCCTCATCCTGGAGCTGGTCTCTG GCGGGGAGCTCTTTGACTTCCTGGCGGAGAAAGAGTCGCTGACGGAGGACGAGGC CACCCAGTTCCTCAAGCAGATCCTGGACGGCGTTCACTACCTGCACTCTAAGCGC ATCGCACACTTTGACCTGAAGCCGGAAAACATCATGCTGCTGGACAAGAACGTGC CCAACCCACGAATCAAGCTCATCGACTTCGGCATCGCGCACAAGATCGAGGCGGG GAACGAGTTCAAGAACATCTTCGGCACCCCGGAGTTTGTGGCCCCAGAGATTGTG AACTATGAGCCGCTGGGCCTGGAGGCGGACATGTGGAGCATCGGTGTCATCACCT ATATCCTCCTGAGCGGTGCATCCCCGTTCCTGGGCGAGACCAAGCAGGAGACGCT CACCAACATCTCAGCCGTGAACTACGACTTCGACGAGGAGTACTTCAGCAACACC AGCGAGCTGGCCAAGGACTTCATTCGCCGGCTGCTCGTCAAAGATCCCAAGCGGA GAATGACCATTGCCCAGAGCCTGGAACATTCCTGGATTAAGGCGATCCGGCGGCG GAACGTGCGTGGTGAGGACAGCGGCCGCAAGCCCGAGCGGCGGCGCCTGAAGACC ACGCGTCTGAAGGAGTACACCATCAAGTCGCACTCCAGCTTGCCGCCCAACAACA GCTACGCCGACTTCGAGCGCTTCTCCAAGGTGCTGGAGGAGGCGGCGGCCGCCGA GGAGGGCCTGCGCGAGCTGCAGCGCAGCCGGCGGCTCTGCCACGAGGACGTGGAG GCGCTGGCCGCCATCTACGAGGAGAAGGAGGCCTGGTACCGCGAGGAGAGCGACA GCCTGGGCCAGGACCTGCGGAGGCTACGGCAGGAGCTGCTCAAGACCGAGGCGCT CAAGCGGCAGGCGCAGGAGGAGGCCAAGGGCGCGCTGCTGGGGACCAGCGGCCTC AAGCGCCGCTTCAGCCGCCTGGAGAACCGCTACGAGGCGCTGGCCAAGCAAGTAG CCTCCGAGATGCGCTTCGTGCAGGACCTCGTGCGCGCCCTGGAGCAGGAGAAGCT GCAGGGCGTGGAGTGCGGGCTGCGCTAGGCGCAGTGGGGTGGGCCAGGCCCCAGG ACAGCCGGAGCTCGGCCTGCGGTGGGGGCGCTTCCTGTGGACGCTGCGCCTCCCA TCGCCCGGGTGCCTGTCCTTGCCCAGCGCCACCAGGCTGGAGGCGGAGTGGGAGG AGCTGGAGCCAGGCCCGTAAGTTCGCAGGCAGGGGTGGGTGTGGGACGGGGCTGC TTCTCTACACAGCCTCTACGCTGGCCTTCACCTTCACCCCTGCATCGTCGGTGAC CCTGGGACCCTCCAGGCAGCGTGGCCTGTGGCACCGTGAGGGTTGGGACCCACCG AGGCGCAGAGGCGGCCCGAATGCAGCCCTGGTTCAGGCCCGGAGGAGGGTTTGCG GGTAGTTGCACGGACAATTCGGCGGGGTGCTGCCTGTTGCTGCCATTAGCCCAGG AGGAGGTCGTGGGACGGGGAGGGTGGGATGGACGGCGGACAGGCAGTCCCCACGC TGCTGGGTGGCGCCGGGCTTGGTGGGGTCTTCCACTGTGTGCCCTTCTCGCCGAG GCCGGTCCCCCGGGTGTGGGGTGCCCTGCTGCGGACTCCTCCGCGAGCCCCATCG

	TCGCGCCTGTGGACGCCTAGGCAAGAGCGGCCCTCTGCAGCCAAGAGAAATAAAA TACTGGCTTCCAGATAAAAAAAAAAAAAAAAAAAAAAAAAA
	SEQ ID NO: 119 454 aa
DAPK3 CG56543-01 DNA Sequence	MSTFRQEDVEDHYEMGEELGSGQFAIVRKCRQKGTGKEYAAKFIKKRRLSSSRRG VSREEIEREVNILREIRHPNIITLHDIFENKTDVVLILELVSGGELFDFLAEKES LTEDEATQFLKQILDGVHYLHSKRIAHFDLKPENIMLLDKNVPNPRIKLIDFGIA HKIEAGNEFKNIFGTPEFVAPEIVNYEPLGLEADMWSIGVITYILLSGASPFLGE TKQETLTNISAVNYDFDEEYFSNTSELAKDFIRRLLVKDPKRRMTIAQSLEHSWI KAIRRRNVRGEDSGRKPERRRLKTTRLKEYTIKSHSSLPPNNSYADFERFSKVLE EAAAAEEGLRELQRSRRLCHEDVEALAAIYEEKEAWYREESDSLGQDLRRLRQEL LKTEALKRQAQEEAKGALLGTSGLKRRFSRLENRYEALAKQVASEMRFVQDLVRA LEQEKLQGVECGLR
	SEQ ID NO: 120 3122 bp
MEIS2B DNA Sequence	GGCACGAGGGCCCTTGGCTACATCGGACCCAGATGACTGCCTCCTCACTTCCTCC CTCCCGATTCCGCCGCGGCCCCCAAAGACTCTCGGGGTGGCCCCTTGTCCGCACC GCTTGGAGGGAGTGTGCTCTGAGTTAAGCTGGTCTCTTCTGGTCCTGGAAAAAAA TGAGTATTGACAAGGTTGCTGGATCTGCGTAGAAAAGAAAGTGCCACTTAATAAA AAATTTAGCCCGGCAGTGGTACCGTCTGCAGAGCTTGCTGCCCTTGGACGTTAGC AGGAAGCCTTCGGGGTGCTGTAATCGGCGGGCAGAGGAGAGGGAGGCCGCGGAAT TAAAAGGAACAAAAGCTAGAGCGCCATGCCAAACGTCCCCGGCAAGACCCAGTTA GGCAGGAGCCGGGAGTGATGGGAAAATGAACTAGAATACGATGAGCTGCCCCATT ACGGCGGGATGGACGGAGTAGGGGTTCCCGCTTCCATGTACGGAGACCCTCACGC GCCGCGGCCGATCCCCCCGGTTCACCACCTGAACCACGGGCCGCCGCTCCACGCC ACACAGCACTACGGCGCGCACGCCCCGCACCCCAATGTCATGCCGGCCAGTATGG GATCCGCTGTCAACGACGCCTTGAAGCGGGACAAGGACGCGATCTATGGGCACCC GTTGTTTCCTCTGTTAGCTCTGGTCTTTGAGAAGTGCGAGCTGGCGACCTGCACT CCCCGGGAACCTGGAGTGGCTGGCGGAGACGTCTGCTCCTCCGACTCCTTCAACG AGGACATCGCGGTCTTCGCCAAGCAGGTTCGCGCCGAAAAGCCACTTTTTTCCTC AAATCCAGAGCTGGACAATTTGATGATACAAGCAATACAAGTACTAAGGTTTCAT CTTTTGGAGTTAGAAAAGGTCCACGAACTGTGCGATAACTTCTGCCACCGATACA TTAGCTGTTTGAAGGGGAAAATGCCCATCGACCTCGTCATTGATGAAAGAGACGG CAGCTCCAAGTCAGATCATGAAGAACTTTCAGGCTCCTCCACAAATCTCGCTGAC CATAACCCTTCTTCTTGGCGAGACCACGATGATGCAACCTCAACCCACTCAGCAG GCACCCCAGGGCCCTCCAGTGGGGGCCATGCTTCCCAGAGCGGAGACAACAGCAG TGAGCAAGGGGATGGTTTAGACAACAGTGTAGCTTCACCTGGTACAGGTGACGAT GATGATCCGGATAAGGACAAAAAACGCCAGAAGAAAAGAGGCATTTTCCCCAAAG TAGCAACAAATATCATGAGAGCATGGCTCTTCCAGCATCTCACACATCCGTACCC TTCCGAAGAGCAGAAGAAACAGTTAGCGCAAGACACAGGACTTACAATTCTCCAA GTAAACAACTGGTTTATTAATGCCAGAAGAAGAATAGTACAGCCCATGATTGACC AGTCAAATCGAGCAGTGAGCCAAGGAGCAGCATATAGTCCAGAGGGTCAGCCCAT GGGGAGCTTTGTGTTGGATGGTCAGCAACACATGGGGATCCGGCCTGCAGGACCT ATGAGTGGAATGGGCATGAATATGGGCATGGATGGGCAATGGCACTACATGTAAC CTTCATCATGTAAAGCAATCGCAAAGCAAGGGGGAAGTTTGCAGAGCATGCCAGG GGACTACGTTTCTCAGGGTGGTCCTATGGGAATGAGTATGGCACAGCCAAGTTAC ACTCCTCCCCAGATGACCCCACACCCTACTCAATTAAGACATGGACCCCCAATGC ATTCATATTTGCCAAGCCATCCCCACCACCCAGCCATGATGATGCACGGAGGACC CCCTACCCACCCTGGAATGACTATGTCAGCACAGAGCCCCACAATGTTAAATTCT GTAGATCCCAATGTTGGCGGACAGGTTATGGACATTCATGCCCAATAGTATAAGG GAACTCAAGGGAAAAGGAAACACACGCAAAAACTATTTTAAGACTTTCTGAACTT TGACCAGATGTTGACACTTAATATGAAATTCCAGACAGCTGTGATTATTTTTTAC TTTTGTCATTTTTCATCAAGCAACAGAGGACCAATGCAACAAGAACACAAATGTG AAATCATGGGCTGACTGAGACAATTCTGTCCATGTAAAGATCCTCTGGAAAAAGA CTCCGAGAGTTATAACTACTGTAGTATAAATATAGGAACTAAGTTAAACTTGTAC ATTTCTGTTGATCACGCCGTTATGTTGCCTCAAATAGTTTTAGAAGAGAAAAAAA

	AATATATCCTTGTTTTCCACACTATGTGTGTTGTTCCCAAAAGAATGACTGTTTT GGTTCATCAGTGAATTCACCATCCAGGAGAGACTGTGGTATATATTTTAAACCTG TTGGGCCAATGAGAAAAGAACCACACTGGAGATCATGATGAACTTTTGGCTGAAC CTCATCACTCGAACTCCAGCTTCAAGAATGTGTTTTCATGCCCGGCCTTTGTTCC TCCATAAATGTGTCCTTTAGTTTCAAACAGATCTTTATAGTTCGTGCTTCATAAG CCAATTCTTATTATTATTTTTGGGGGACTCTTCTTCAAAGAGCTTGCCAATGAAG ATTTAAAGACAGAGCAGGAGCTTCTTCCAGGAGTTCTGAGCCTTGGTTGTGGACA AAACAATCTTAAGTTGGGCAGCTTTCCTCAACACAAAAAAAGTTATTAATGGTCA TTGAACCATAACTAGGACTTTATCAGAAACTCAAAGCTTGGGGGATAAAAAGGAG CAAGAGAATACTGTAACAAACTTCGTACAGAGTTCGGTCTATTAATTGTTTCATG TTAGATATTCTATGTGTTTACCTCAATTGAAAAAAAAAAAGAATGTTTTTGCTAGT ATCAGATCTGCTGTGGAATTGGTATTGTATGTCCATGAATTCTTCTTTTCTCAGC ACGTGTTCCTCACTAGAAGAAAATGCTGTTACCTTTAAGCTTTGTCAAATTTACA TTAAAATACTTGTATGAGGACTGTGACGTTATGTTAAAAAAAAAAAGGTGTTAAGT CACAAAAAGCGGTAATAAATATTTCATTTTTGAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
	SEQ ID NO: 121 381 aa
MEIS2B Protein Sequence	MDGVGVPASMYGDPHAPRPIPPVHHLNHGPPLHATQHYGAHAPHPNVMPASMGSA VNDALKRDKDAIYGHPLFPLLALVFEKCELATCTPREPGVAGGDVCSSDSFNEDI AVFAKQVRAEKPLFSSNPELDNLMIQAIQVLRFHLLELEKVHELCDNFCHRYISC LKGKMPIDLVIDERDGSSKSDHEELSGSSTNLADHNPSSWRDHDDATSTHSAGTP GPSSGGHASQSGDNSSEQGDGLDNSVASPGTGDDDDPDKDKKRQKKRGIFPKVAT NIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPMIDQSN RAVSQGAAYSPEGQPMGSFVLDGQQHMGIRPAGPMSGMGMNMGMDGQWHYM

Example F: TRAF5 Interactions

Novel associations of TRAF5 proteins and TRAF5 interacting proteins (TRAF-IP), and the nucleic acids that encode them, are described, as are various diseases or pathologies associated with TRAF5 and TRAF-IP protein complexes (TRAF5:TRAF-IP). The TRAF-IP proteins, polypeptides and their cognate nucleic acids were identified by Curagen Corporation in certain cases. The TRAF5:TRAF-IP protein complexes, and any variants thereof, are suitable as targets for an antibody therapeutic and targets for small molecule drugs. As such the current invention embodies the use of recombinantly expressed and/or endogenously expressed TRAF5:TRAF-IP protein complexes in various screens to identify such therapeutic antibodies and/or therapeutic small molecules.

Discovery Method:

TRAF-IPs were identified using PathCalling™ Technology (CuraGen Corporation). The sequences was derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were sequenced. In silico prediction was based on sequences available in Curagen Corporation's proprietary sequence databases or in the public human

sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods described in Example E.

Description of the Interaction:

Lymphotoxin beta receptor (LTβR), a member of the TNF-receptor superfamily is required for the development and organization of the lymphoid tissue. Besides its role in the immune system, it is also required for the development of diabetes in nonobese diabetic mice. It is expressed in a variety of normal and transformed cells but absent in T and B lymphocytes. In vitro it can induce growth stimulation of fibroblasts and is associated with a variety of inflammatory disorders. LTβR induces signaling via its interaction with TRAF proteins, TRAF2, 3 and 5. It activates NFkB signaling via the activation of NFkB inducing kinase.

A primary function of TRAF5 is protection of cells against cell death via activation of NFkB. One of the two TRAF5 proteins described in the interaction (TRAF5 SV) is a novel splice variant that lacks amino acid residues 232-263 present in the TRAF5 protein available from the public data base. The other TRAF5 (novel TRAF5) described in the interaction has extra 11 amino acids inserted between residues 125-126 of the public gene. In the PathCalling™ interaction, full-length TRAF5 SV and residues 280-569 of the novel TRAF5 were used.

As shown in Table F1, both of the TRAF5 protein constructs interacted with LTβR receptor as expected. This interaction is likely via their TRAF domains. In addition, both the TRAF5 protein constructs interacted with the human ubiquitinating enzyme (AAT62352), which is highly homologous to the ubiquitin-like protein SUMO-1 conjugating enzyme. This is a novel interaction as no other TRAF protein has been shown to interact with this enzyme. The interaction is significant because it is known from the literature that TRAF6 is multi-ubiquitinated via Ubc13-Uev1A. Such ubiquitination does not target TRAF6 for degradation via the proteasomal pathway and is essential for TRAF6-dependent signaling. Based on the interaction between TRAF5 and the ubiquitinating enzyme, it can be hypothesized that TRAF5 is modified via an ubiquitin-like protein and this modification is important in TRAF5-mediated signaling.

A second interesting interactor of the novel TRAF5 protein is with the Activatory receptor protein (CG124499-01), which belongs to the Immunoglobulin-superfamily of receptors and homologous to the PILRα protein. These receptors bind to their inhibitory

counterpart on the cell surface like PILRα binds PILRβ and counteract each others activities. Interaction of TRAF5 with the activating receptor suggests a novel mechanism of signaling from these family of receptors. Another interactor of TRAF5 is a protein that binds activated STAT3 (PIAS3) and is a negative regulator of JAK/STAT signaling similar to the SOCS proteins. Sequestration of PIAS3 is therefore a likely mechanism of activating the JAK/STAT pathway. The activatory receptor (CG124499-01) may influence JAK/STAT pathway by removing PIAS3 from STAT3 protein utilizing TRAF5. In a broader perspective, TRAF proteins may play an interesting role in receptor cross-talk between different cell surface receptors.

Uses of the Compositions of the Invention:

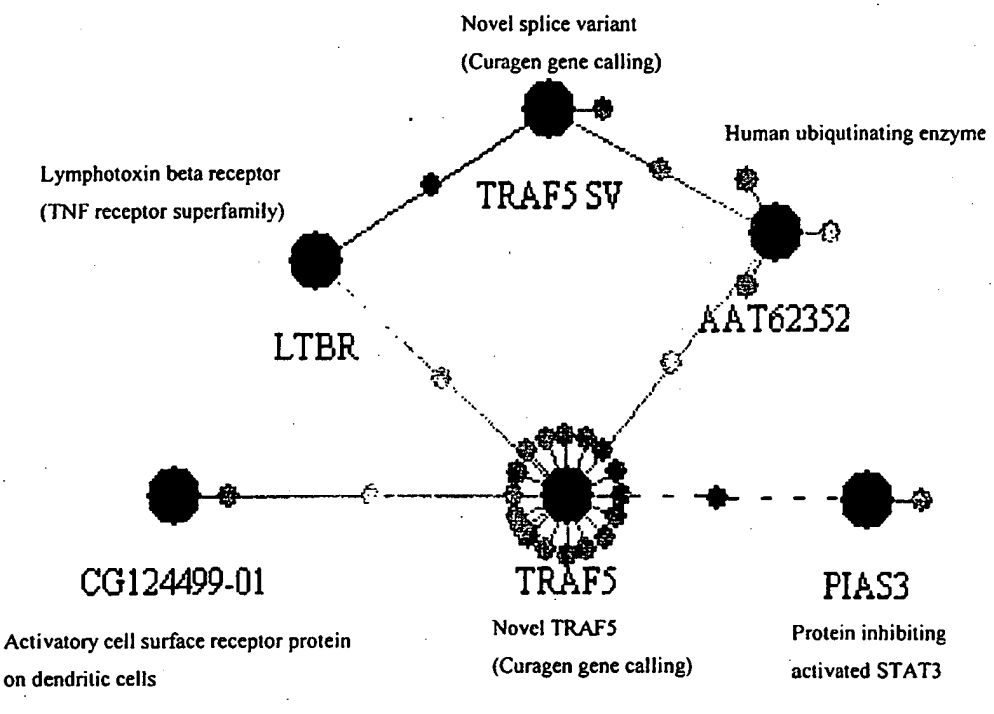
The interactors of TRAF5 and their relevance to TRAF5 and in general to TRAF signaling provides opportunities to develop tools against various pathologic situations in which signaling through TRAF proteins are involved, and against the receptors that regulate the signaling pathways. Therefore, the TRAF5:TRAF-IP complexes of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These tools include identification of TRAF5 - TRAF-IP associations for a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein is to be assessed. Potential therapeutic applications for the TRAF5:TRAF-IP complexes are the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

The nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: cancer; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; colon cancer, leukemia AIDS; metabolic disorders including diabetes and obesity; pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the

invention for use in therapeutic or diagnostic methods. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

The identified DAPK3 protein interactions and their corresponding biological pathways are shown in Table F1.

Table F1. Diagram of Protein Interactions and/or Biological Pathway:



The nucleotide and polypeptide sequences of the identified DAPK-IPs are provided in Table F2.

Table F2. Nucleic Acid and Amino Acid Sequence Analysis of TRAF5 Interacting Proteins

	SEQ ID NO: 122	1968 bp	
LTβR GDB: 1230195 DNA Sequence	ATGCTCCTGCCTTGGGCCACCTCTGCCCCCGGCCTGGCCTGGGGGCCTCTGGTGCTGGGCC TCTTCGGGCTCCTGGCAGCATCGCAGCCCCAGGCGGTGCCTCCATATGCGTCGGAGAACCA GACCTGCAGGGACCAGGAAAAGGAATACTATGAGCCCCAGCACCGCATCTGCTGCTCCCGC TGCCCGCCAGGCACCTATGTCTCAGCTAAATGTAGCCGCATCCGGGACACAGTTTGTGCCA CATGTGCCGAGAATTCCTACAACGAGCACTGGAACTACCTGACCATCTGCCAGCTGTGCCG CCCCTGTGACCCAGTGATGGGCCTCGAGGAGATTGCCCCCTGCACAAGCAAACGGAAGACC CAGTGCCGCTGCCAGCCGGGAATGTTCTGTGCTGCCTGGGCCCTCGAGTGTACACACTGCG AGCTACTTTCTGACTGCCCGCCTGGCACTGAAGCCGAGCTCAAAGATGAAGTTGGGAAGGG		

	TAACAACCACTGCGTCCCCTGCAAGGCAGGGCACTTCCAGAATACCTCCTCCCCCAGCGCC CGCTGCCAGCCCCACACCAGGTGTGAGAACCAAGGTCTGGTGGAGGCAGCTCCAGGCACTG CCCAGTCCGACACAACCTGCAAAAATCCATTAGAGCCACTGCCCCCAGAGATGTCAGGAAC CATGCTGATGCTGGCCGTTCTGCTGCCACTGGCCTTCTTTCTGCTCCTTGCCACCGTCTTC TCCTGCATCTGGAAGAGCCACCCTTCTCTCTGCAGGAAACTGGGATCGCTGCTCAAGAGGC GTCCGCAGGGAGAGGGACCCAATCCTGTAGCTGGAAGCTGGGAGCCTCCGAAGGCCCATCC ATACTTCCCTGACTTGGTACAGCCACTGCTACCCATTTCTGGAGATGTTTCCCCAGTATCC ACTGGGCTCCCCGCAGCCCCAGTTTTGGAGGCAGGGGTGCCGCAACAGCAGAGTCCTCTGG ACCTGACCAGGGAGCCGCAGTTGGAACCCGGGGAGCAGAGCCAGGTGGCCCACGGTACCAA TGGCATTCATGTCACCGGCGGGTCTATGACTATCACTGGCAACATCTACATCTACAATGGA CCAGTACTGGGGGGACCACCGGGTCCTGGAGACCTCCCAGCTACCCCCGAACCTCCATACC CCATTCCCGAAGAGGGGGACCCTGGCCCTCCCGGGCTCTCTACACCCCACCAGGAAGATGG CAAGGCTTGGCACCTAGCGGAGACAGAGCACTGTGGTGCCACACCCTCTAACAGGGGCCCA AGGAACCAATTTATCACCCATGACTGACGGAGTCTGAGAAAAGGCAGAAGAAGGGGGGCAC AAGGGCACTTTCTCCCTTGAGGCTGCCCTGCCCACGTGGGATTCACAGGGGCCTGAGTAGG GCCCGGGGAAGCAGAGCCCTAAGGGATTAAGGCTCAGACACCTCTGAGAGCAGGTGGGCAC TGGCTGGGTACGGTGCCCTCCACAGGACTCTCCCTACTGCCTGAGCAAACCTGAGGCCTCC CGGCAGACCCACCCACCCCCTGGGGCTGCTCAGCCTCAGGCACGGACAGGGCACATGATAC CAACTGCTGCCCACTACGGCACGCCGCACCGGAGCACGGCACCGAGGGAGCCGCCACACGG TCACCTGCAAGGACGTCACGGGCCCCTCTAAAGGATTCGTGGTGCTCATCCCCAAGCTTCA GAGACCCTTTGGGGTTCCACACTTCACGTGGACTGAGGTAGACCCTGCATGAAGATGAAAT TATAGGGAGGACGCTCCTTCCCTCCCCTCCTAGAGGAGAGGAAAGGGAGTCATTAACAACT AGGGGGTTGGGTAGGATTCCTAGGTATGGGGAAGAGTTTTGGAAGGGGAGGAAAATGGCAA GTGTATTTATATTGTAACCACATGCAAATAAAAAGAATGGGACCTAAACTCGTGCCGCTCG TGCCGAATTCCTGCAG
	SEQ ID NO: 435 aa
LTβR GDB: 1230195 Protein Sequence	MLLPWATSAPGLAWGPLVLGLFGLLAASQPQAVPPYASENQTCRDQEKEYYEPQHRICCSR CPPGTYVSAKCSRIRDTVCATCAENSYNEHWNYLTICQLCRPCDPVMGLEEIAPCTSKRKT QCRCQPGMFCAAWALECTHCELLSDCPPGTEAELKDEVGKGNNHCVPCKAGHFQNTSSPSA RCQPHTRCENQGLVEAAPGTAQSDTTCKNPLEPLPPEMSGTMLMLAVLLPLAFFLLLATVF SCIWKSHPSLCRKLGSLLKRRPQGEGPNPVAGSWEPPKAHPYFPDLVQPLLPISGDVSPVS TGLPAAPVLEAGVPQQQSPLDLTREPQLEPGEQSQVAHGTNGIHVTGGSMTITGNIYIYNG PVLGGPPGPGDLPATPEPPYPIPEEGDPGPPGLSTPHQEDGKAWHLAETEHCGATPSNRGP RNQFITHD
	SEQ ID NO: 123 1148 bp
Human deubiquitinating enzyme DNA Sequence	AGCGGGCTCCGGAGGGAAGTCCCGAGACAAAGGGAAACGCCGCCGCCGCCGCCCCGCTCGG TCCTCCACCTGTCCGCTACGCTCGCCAGGGCTGCGGCCGCCCGAGGGACTTTGAACATGTC GGGGATCGCCCTCAGCAGACTCGCCCAGGAGAGGAAAGCATGGAGGAAAGACCACCCATTT GGTTTCGTGGCTGTCCCAACAAAAAATCCCGATGGCACGATGAACCTCATGAACTGGGAGT GCGCCATTCCAGGAAAGAAAGGGACTCCGTGGGAAGGAGGCTTGTTTAAACTACGGATGCT TTTCAAAGATGATTATCCATCTTCGCCACCAAAATGTAAATTCGAACCACCATTATTTCAC CCGAATGTGTACCCTTCGGGGACAGTGTGCCTGTCCATCTTAGAGGAGGACAAGGACTGGA GGCCAGCCATCACAATCAAACAGATCCTATTAGGAATACAGGAACTTCTAAATGAACCAAA TATCCAAGACCCAGCTCAAGCAGAGGCCTACACGATTTACTGCCAAAACAGAGTGGAGTAC GAGAAAAGGGTCCGAGCACAAGCCAAGAAGTTTGCGCCCTCATAAGCAGCGACCTTGTGGC ATCGTCAAAAGGAAGGGATTGGTTTGGCAAGAACTTGTTTACAACATTTTTGCAAATCTAA AGTTGCTCCATACAATGACTAGTCACCTGGGGGGGTTGGGCGGGCGCCATCTTCCATTGCC GCCGCGGGTGTGCGGTCTCGATTCGCTGAATTGCCCGTTTCCATACAGGGTCTCTTCCTTC GGTCTTTTGTATTTTTGATTGTTATGTAAAACTCGCTTTTATTTTAATATTGATGTCAGTA TTTCAACTGCTGTAAAATTATAAACTTTTATACTTGGGTAAGTCCCCCAGGGGCGAGTTCC TCGCTCTGGGATGCAGGCATGCTTCTCACCGTGCAGAGCTGCACTTGGCCTCAGCTGGCTG TATGGAAATGCACCCTCCCTCCTGCCGCTCCTCTCTAGAACCTTCTAGAACCTGGGCTGTG CTGCTTTTGAGCCTCAGACCCCAGGTCAGCATCTCGGTTCTGCGCCACTTCCTTTGTGTTT ATATGGCGTTTTGTCTGTGTTGCTGTTTAGAGTAAATAAACTGTTTATAT
	SEQ ID NO: 124 158 aa
Human deubiquitinating enzyme	MSGIALSRLAQERKAWRKDHPFGFVAVPTKNPDGTMNLMNWECAIPGKKGTPWEGGLFKLR MLFKDDYPSSPPKCKFEPPLFHPNVYPSGTVCLSILEEDKDWRPAITIKQILLGIQELLNE PNIQDPAQAEAYTIYCQNRVEYEKRVRAQAKKFAPS

Protein Sequence	
	SEQ ID NO: 125 2808 bp
Protein inhibiting activated STAT3 (PIAS3) DNA Sequence	GGCGGAGCTGGGCGAATTAAAGCACATGGTGATGAGTTTCCGGGTGTCTGAGCTCCAGGTG CTTCTTGGCTTTGCTGGCCGGAACAAGAGTGGACGGAAGCACGAGCTCCTGGCCAAGGCTC TGCACCTCCTGAAGTCCAGCTGTGCCCCTAGTGTCCAGATGAAGATCAAAGAGCTTTACCG ACGACGCTTTCCCCGGAAGACCCTGGGGCCCTCTGATCTCTCCCTTCTCTCTTTGCCCCCT GGCACCTCTCCTGTAGGCTCCCCTGGTCCTCTAGCTCCCATTCCCCCAACGCTGTTGGCCC CTGGCACCCTGCTGGGCCCCAAGCGTGAGGTGGACATGCACCCCCCTCTGCCCCAGCCTGT GCACCCTGATGTCACCATGAAACCATTGCCCTTCTATGAAGTCTATGGGGAGCTCATCCGG CCCACCACCCTTGCATCCACTTCTAGCCAGCGGTTTGAGGAAGCGCACTTTACCTTTGCCC TCACACCCCAGCAAGTGCAGCAGATTCTTACATCCAGAGAGGTTCTGCCAGGAGCCAAATG TGATTATACCATACAGGTGCAGCTAAGGTTCTGTCTCTGTGAGACCAGCTGCCCCCAGGAA GATTATTTTCCCCCCAACCTCTTTGTCAAGGTCAATGGGAAACTGTGCCCCCTGCCGGGTT ACCTTCCCCCAACCAAGAATGGGGCCGAGCCCAAGAGGCCCAGCCGCCCCATCAACATCAC ACCCCTGGCTCGACTCTCAGCCACTGTTCCCAACACCATTGTGGTCAATTGGTCATCTGAG TTCGGACGGAATTACTCCTTGTCTGTGTACCTGGTGAGGCAGTTGACTGCAGGAACCCTTC TACAAAAACTCAGAGCAAAGGGTATCCGGAACCCAGACCACTCGCGGGCACTGATCAAGGA GAAATTGACTGCTGACCCTGACAGTGAGGTGGCCACTACAAGTCTCCGGGTGTCACTCATG TGCCCGCTAGGGAAGATGCGCCTGACTGTCCCTTGTCGTGCCCTCACCTGCGCCCACCTGC AGAGCTTCGATGCTGCCCTTTATCTACAGATGAATGAGAAGAAGCCTACATGGACATGTCC TGTGTGTGACAAGAAGGCTCCCTATGAATCTCTTATCATTGATGGTTTATTTATGGAGATT CTTAGTTCCTGTTCAGATTGTGATGAGATCCAATTCATGGAAGATGGATCCTGGTGCCCAA TGAAACCCAAGAAGGAGGCATCTGAGGTTTGCCCCCCGCCAGGGTATGGGCTGGATGGCCT CCAGTACAGCCCAGTCCAGGGGGGAGATCCATCAGAGAATAAGAAGAAGGTCGAAGTTATT GACTTGACAATAGAAAGCTCATCAGATGAGGAGGATCTGCCCCCTACCAAGAAGCACTGTT CTGTCACCTCAGCTGCCATCCCGGCCCTACCTGGAAGCAAAGGAGTCCTGACATCTGGCCA CCAGCCATCCTCGGTGCTAAGGAGCCCTGCTATGGGCACGTTGGGTGGGGATTTCCTGTCC AGTCTCCCACTACATGAGTACCCACCTGCCTTCCCACTGGGAGCCGACATCCAAGGTTTAG ATTTATTTTCATTTCTTCAGACAGAGAGTCAGCACTATGGCCCCTCTGTCATCACCTCACT AGATGAACAGGATGCCCTTGGCCACTTCTTCCAGTACCGAGGGACCCCTTCTCACTTTCTG GGCCCACTGGCCCCCACGCTGGGGAGCTCCCACTGCAGCGCCACTCCGGCGCCCCCTCCTG GCCGTGTCAGCAGCATTGTGGCCCCTGGGGGGGCCTTGAGGGAGGGGCATGGAGGACCCCT GCCCTCAGGTCCCTCTTTGACTGGCTGTCGGTCAGACATCATTTCCCTGGACTGAGTTCCC TGGATTATGGAAACTTCGCTGTCCCCCAACACTGAGCAAGTATGCTGTGGAGTCCCAACCC CAGCTACTCTGATCCCTCTGGGGGCTCTGGCCAAGGGCCAGACAGACCTTCACAGATGCCT ACTTTTGGCCTCATCTCTGCCTGACAAGGCCAGCACCCAAAGGGTTAATATTTAACCTCTT TTTAAGGACACTGGGGTCTGTTTCTGGAAATGTTCTTTAGATGGTGGCACATTCCTTTGGG TATGTTAACCTAGGCAGTGGGAGGCAAATGGGATGGTATGTGAGCTAGGAGAAGGGCTGAA CCCTCAGCCTTGACTATGTCTAGAGCCTCTTGGGGAAGGGGCACCTCTCTTGAACCCCAAA TGCTCTCTCTTCTTATTACCCAAACCCATGGCTCTATTTCTTCTTCACATCCATTGTCTCT TCATGTCTATTCCATTCCCTTCGGCCAAACAGACAGGTGGAAAAACTGAGACAGGCAGTTT CAGAGATGGACAGAGAACTTTATTTTGGATTGTGGATGTGGACTTTTTTGTACATAAATAA GAAAAACCAAAATACTCCAAAGATGACTTCCCCTGCCTCCTACTCCAGTATGACAGAGGAG GATGTAAGGCCTTAGCCATGATCTGCAGGGGTCTGGGAGTCAGGCCCGGCCTATTGCTTGG GTCTCTCTCTATTTATATATCTAAGTTCACAGTGTTTCTTATTCCCCCTAAGCTTCTAGAG GCTCATGGCCCTGTAGTTAGGCCTGGCTCATTCTGCACCTTTCCAGGGAGGTGGAAGGACC CTGTGCCCTCCTTCCCAATCTTCTTTTTCAGGCTCGCCAAGGCCTAGGACCTATGTTGTAA TTTTACTTTTTATTTCTAAAGTTGTAGTGAAGCTCTCACCCATAATAAAGGTTGTGAATGT TC
	SEQ ID NO: 126 619 aa
Protein inhibiting activated STAT3 (PIAS3) Protein Sequence	MVMSFRVSELQVLLGFAGRNKSGRKHELLAKALHLLKSSCAPSVQMKIKELYRRRFPRKTL GPSDLSLLSLPPGTSPVGSPGPLAPIPPTLLAPGTLLGPKREVDMHPPLPQPVHPDVTMKP LPFYEVYGELIRPTTLASTSSQRFEEAHFTFALTPQQVQQILTSREVLPGAKCDYTIQVQL RFCLCETSCPQEDYFPPNLFVKVNGKLCPLPGYLPPTKNGAEPKRPSRPINITPLARLSAT VPNTIVVNWSSEFGRNYSLSVYLVRQLTAGTLLQKLRAKGIRNPDHSRALIKEKLTADPDS EVATTSLRVSLMCPLGKMRLTVPCRALTCAHLQSFDAALYLQMNEKKPTWTCPVCDKKAPY ESLIIDGLFMEILSSCSDCDEIQFMEDGSWCPMKPKKEASEVCPPPGYGLDGLQYSPVQGG DPSENKKKVEVIDLTIESSSDEEDLPPTKKHCSVTSAAIPALPGSKGVLTSGHQPSSVLRS

	PAMGTLGGDFLSSLPLHEYPPAFPLGADIQGLDLFSFLQTESQHYGPSVITSLDEQDALGH FFQYRGTPSHFLGPLAPTLGSSHCSATPAPPPGRVSSIVAPGGALREGHGGPLPSGPSLTG CRSDIISLD
	SEQ ID NO: 127 | 1450 bp |
Activatory cell surface receptor protein CG124499-01 DNA Sequence	CCACGCGTCCGGCTTCTTTGGGGGTGAAGAGATTGGGGAGGAATCTCCACCCCTGGGAGGC AGAAGCCAGGCATAGCGCGCTGGCTAGGACTCCAGTACCGTGAAGGGAGGCAGTGAGAGCA GACATCTGTGCCTCATTCCTGATCTCAAGGGGAAAGCAAGAACAAGGGAGGCTTCCTCAGG ATCTCGAACCTGCGGAAGGAGGACCAGTCTGTGTACTTCTGCCAAGTCCAGCTGGACATAC AGATCAGGGAGGCTGTCGTGGCAGTCCATCAAGGGGACCCACCTCACCATCACCCAGGCCC TCAGGCAGCCCCTCCACAGGGCCCCTCTCCTGCCTGGACAGCTCTGCTGGTCTCCCCGTCC CCTGGAGAAGAACAAGGCCATGGGTCGGCCCCTGCTGCTGCCCCTGCTGCTCCTGCTGCAG CCGCCAGCATTTCTGCAGCCTGGTGGCTCCACAGGATCTGGTCCAAGCTACCTTTATGGGG TCACTCAACCAAAACACCTCTCAGCCTCCATGGGTGGCTCTGTGGAAATCCCCTTCTCCTT CTATTACCCCTGGGAGTTAGCCATAGTTCCCAACGTGAGAATATCCTGGAGACGGGGCCAC TTCCACGGGCAGTCCTTCTACAGCACAAGGCCGCCTTCCATTCACAAGGATTATGTGAACC GGCTCTTTCTGAACTGGACAGAGGGTCAGGAGAGCGGCTTCCTCAGGATCTCAAACCTGCG GAAGGAGGACCAGTCTGTGTATTTCTGCCGAGTCGAGCTGGACACCCGGAGATCAGGGAGG CAGCAGTTGCAGTCCATCAAGGGGACCAAACTCACCATCACCCAGGCTGTCACAACCACCA CCACCTGGAGGCCCAGCAGCACAACCACCATAGCCGGCCTCAGGGTCACAGAAAGCAAAGG GCACTCAGAATCATGGCACCTAAGTCTGGACACTGCCATCAGGGTTGCATTGGCTGTCGCT GTGCTCAAAACTGTCATTTTGGGACTGCTGTGCCTCCTCCTCCTGTGGTGGAGGAGAAGGA AAGGTAGCAGGGCGCCAAGCAGTGACTTCTGACCAACAGAGTGTGGGGAGAAGGGATGTGT ATTAGCCCCGGAGGACGTGATGTGAGACCCGCTTGTGAGTCCTCCACACTCGTTCCCCATT GGCAAGATACATGGAGAGCACCCTGAGGACCTTTAAAAGGCAAAGCCGCAAGGCAGAAGGA GGCTGGGTCCCTGAATCACCGACTGGAGGAGAGTTACCTACAAGAGCCTTCATCCAGGAGC ATCCACACTGCAATGATATAGGAATGAGGTCTGAACTCCACTGAATTAAACCACTGGCATT TGGGGGCTGTTTATTATAGCAGTGCAAAGAGTTCCTTTATCCTCCCCAAGGATGGAAAAAT ACAATTTATTTTGCTTACCATAAAAAAAAAAAAAAAAAAAAAAAAA
	SEQ ID NO: 128 | 227 aa |
Activatory cell surface receptor protein CG124499-01 Protein Sequence	MGRPLLLPLLLLLQPPAFLQPGGSTGSGPSYLYGVTQPKHLSASMGGSVEIPFSFYYPWEL AIVPNVRISWRRGHFHGQSFYSTRPPSIHKDYVNRLFLNWTEGQESGFLRISNLRKEDQSV YFCRVELDTRRSGRQQLQSIKGTKLTITQAVTTTTTWRPSSTTTIAGLRVTESKGHSESWH LSLDTAIRVALAVAVLKTVILGLLCLLLLWWRRRKGSRAPSSDF
	SEQ ID NO: 129 | 1581 bp |
TRAF5 splice variant TRAF5_SV DNA Sequence	ATGGCTTATTCAGAAGAGCATAAAGGTATGCCCTGTGGTTTCATCCGCCAGAATTCCGGCA ACTCCATTTCCTTGGACTTTGAGCCCAGTATAGAGTACCAGTTTGTGGAGCGGTTGGAAGA GCGCTACAAATGTGCCTTCTGCCACTCGGTGCTTCACAACCCCCACCAGACAGGATGTGGG CACCGCTTCTGCCAGCACTGCATCCTGTCCCTGAGAGAATTAAACACAGTGCCAATCTGCC CTGTAGATAAAGAGGTCATCAAATCTCAGGAGGTTTTTAAAGACAATTGTTGCAAAAGAGA AGTCCTCAACTTATATGTATATTGCAGCAATGCTCCTGGATGTAATGCCAAGGTTATTCTG GGCCGGTACCAGGATCACCTTCAGCAGTGCTTATTTCAACCTGTGCAGTGTTCTAATGAGA AGTGCCGGGAGCCAGTCCTACGGAAAGACCTGAAAGAGCATTTGAGTGCATCCTGTCAGTT TCGAAAGGAAAAATGCCTTTATTGCAAAAAGGATGTGGTAGTCATCAATCTACAGAATCAT GAGGAAAACTTGTGTCCTGAATACCCAGTATTTTGTCCCAACAATTGTGCGAAGATTATTC TAAAAACTGAGGTAGATGAACACCTGGCTGTATGTCCTGAAGCTGAGCAAGACTGTCCTTT TAAGCACTATGGCTGTGCTGTAACGATTTCTGACTTACACAAGAGCCTAGAACAGAAAGAA AGTAAAATCCAGCAGCTAGCAGAAACTATAAAGAAACTTGAAAAGGAGTTCAAGCAGTTTG CACAGTTGTTTGGCAAAAATGGAAGCTTCCTCCCAAACATCCAGGTTTTTGCCAGTCACAT TGACAAGTCAGCTTGGCTAGAAGCTCAAGTGCATCAATTATTACAAATGGTTAACCAGCAA CAAAATAAATTTGACCTGAGACCTTTGATGGAAGCAGTTGATACAGTGAAACAGAAAATTA CCCTGCTAGAAAACAATGATCAAAGATTAGCCGTTTTAGAAGAGGAAACTAACAAACATGA TACCCACATTAATATTCATAAAGCACAGCTGAGTAAAAATGAAGAGCGATTTAAACTGCTG GAGGGTACTTGCTATAATGGAAAGCTCATTTGGAAGGTGACAGATTACAAGATGAAGAAGA GAGAGGCGGTGGATGGGCACACAGTGTCCATCTTCAGCCAGTCCTTCTACACCAGCCGCTG TGGCTACCGGCTCTGTGCTAGAGCATACCTGAATGGGGATGGGTCAGGGAGGGGGTCACAC CTGTCCCTATACTTTGTGGTCATGCGAGGAGAGTTTGACTCACTGTTGCAGTGGCCATTCA

	GGCAGAGGGTGACCCTGATGCTTCTGGACCAGAGTGGCAAAAAGAACATTATGGAGACCTT CAAACCTGACCCCAATAGCAGCAGCTTTAAAAGACCTGATGGGGAGATGAACATTGCATCT GGCTGTCCCCGCTTTGTGGCTCATTCTGTTTTGGAGAATGCCAAGAACGCCTACATTAAAG ATGACACTCTGTTCTTGAAAGTGGCCGTGGACTTAACTGACCTGGAGGATCTCTAG
	SEQ ID NO: 130 526 aa
TRAF5 splice variant TRAF5_SV Protein Sequence	MAYSEEHKGMPCGFIRQNSGNSISLDFEPSIEYQFVERLEERYKCAFCHSVLHNPHQTGCG HRFCQHCILSLRELNTVPICPVDKEVIKSQEVFKDNCCKREVLNLYVYCSNAPGCNAKVIL GRYQDHLQQCLFQPVQCSNEKCREPVLRKDLKEHLSASCQFRKEKCLYCKKDVVVINLQNH EENLCPEYPVFCPNNCAKIILKTEVDEHLAVCPEAEQDCPFKHYGCAVTISDLHKSLEQKE SKIQQLAETIKKLEKEFKQFAQLFGKNGSFLPNIQVFASHIDKSAWLEAQVHQLLQMVNQQ QNKFDLRPLMEAVDTVKQKITLLENNDQRLAVLEEETNKHDTHINIHKAQLSKNEERFKLL EGTCYNGKLIWKVTDYKMKKREAVDGHTVSIFSQSFYTSRCGYRLCARAYLNGDGSGRGSH LSLYFVVMRGEFDSLLQWPFRQRVTLMLLDQSGKKNIMETFKPDPNSSSFKRPDGEMNIAS GCPRFVAHSVLENAKNAYIKDDTLFLKVAVDLTDLEDL
	SEQ ID NO: 131 1707 bp
TRAF5 DNA Sequence	ATGGCTTATTCAGAAGAGCATAAAGGTATGCCCTGTGGTTTCATCCGCCAGAATTCCGGCA ACTCCATTTCCTTGGACTTTGAGCCCAGTATAGAGTACCAGTTTGTGGAGCGGTTGGAAGA GCGCTACAAATGTGCCTTCTGCCACTCGGTGCTTCACAACCCCCACCAGACAGGATGTGGG CACCGCTTCTGCCAGCACTGCATCCTGTCCCTGAGAGAATTAAACACAGTGCCAATCTGCC CTGTAGATAAAGAGGTCATCAAATCTCAGGAGGTTTTTAAAGACAATTGTTGCAAAAGAGA AGTCCTCAACTTATATGTATATTGCAGCAATGCTCCTGGATGTAATGCCAAGGTTATTCTG GGCCGGTACCAGCAGGTCCCACTGGCCTGTTGTTATCTGTTGCAGGATCACCTTCAGCAGT GCTTATTTCAACCTGTGCAGTGTTCTAATGAGAAGTGCCGGGAGCCAGTCCTACGGAAAGA CCTGAAAGAGCATTTGAGTGCATCCTGTCAGTTTCGAAAGGAAAAATGCCTTTATTGCAAA AAGGATGTGGTAGTCATCAATCTACAGAATCATGAGGAAAACTTGTGTCCTGAATACCCAG TATTTTGTCCCAACAATTGTGCGAAGATTATTCTAAAAACTGAGGTAGATGAACACCTGGC TGTATGTCCTGAAGCTGAGCAAGACTGTCCTTTTAAGCACTATGGCTGTGCTGTAACGGAT AAACGGAGGAACCTGCAGCAACATGAGCATTCAGCCTTACGGGAGCACATGCGTTTGGTTT TAGAAAAGAATGTCCAATTAGAAGAACAGATTTCTGACTTACACAAGAGCCTAGAACAGAA AGAAAGTAAAATCCAGCAGCTAGCAGAAACTATAAAGAAACTTGAAAAGGAGTTCAAGCAG TTTGCACAGTTGTTTGGCAAAAATGGAAGCTTCCTCCCAAACATCCAGGTTTTTGCCAGTC ACATTGACAAGTCAGCTTGGCTAGAAGCTCAAGTGCATCAATTATTACAAATGGTTAACCA GCAACAAAATAAATTTGACCTGAGACCTTTGATGGAAGCAGTTGATACAGTGAAACAGAAA ATTACCCTGCTAGAAAACAATGATCAAAGATTAGCCGTTTTAGAAGAGGAAACTAACAAAC ATGATACCCACATTAATATTCATAAAGCACAGCTGAGTAAAAATGAAGAGCGATTTAAACT GCTGGAGGGTACTTGCTATAATGGAAAGCTCATTTGGAAGGTGACAGATTACAAGATGAAG AAGAGAGAGGCGGTGGATGGGCACACAGTGTCCATCTTCAGCCAGTCCTTCTACACCAGCC GCTGTGGCTACCGGCTCTGTGCTAGAGCATACCTGAATGGGGATGGGTCAGGGAGGGGGTC ACACCTGTCCCTATACTTTGTGGTCATGCGAGGAGAGTTTGACTCACTGTTGCAGTGGCCA TTCAGGCAGAGGGTGACCCTGATGCTTCTGGACCAGAGTGGCAAAAAGAACATTATGGAGA CCTTCAAACCTGACCCCAATAGCAGCAGCTTTAAAAGACCTGATGGGGAGATGAACATTGC ATCTGGCTGTCCCCGCTTTGTGGCTCATTCTGTTTTGGAGAATGCCAAGAACGCCTACATT AAAGATGACACTCTGTTCTTGAAAGTGGCCGTGGACTTAACTGACCTGGAGGATCTCTAG
	SEQ ID NO: 132 568 aa
TRAF5 Protein Sequence	MAYSEEHKGMPCGFIRQNSGNSISLDFEPSIEYQFVERLEERYKCAFCHSVLHNPHQTGCG HRFCQHCILSLRELNTVPICPVDKEVIKSQEVFKDNCCKREVLNLYVYCSNAPGCNAKVIL GRYQQVPLACCYLLQDHLQQCLFQPVQCSNEKCREPVLRKDLKEHLSASCQFRKEKCLYCK KDVVVINLQNHEENLCPEYPVFCPNNCAKIILKTEVDEHLAVCPEAEQDCPFKHYGCAVTD KRRNLQQHEHSALREHMRLVLEKNVQLEEQISDLHKSLEQKESKIQQLAETIKKLEKEFKQ FAQLFGKNGSFLPNIQVFASHIDKSAWLEAQVHQLLQMVNQQQNKFDLRPLMEAVDTVKQK ITLLENNDQRLAVLEEETNKHDTHINIHKAQLSKNEERFKLLEGTCYNGKLIWKVTDYKMK KREAVDGHTVSIFSQSFYTSRCGYRLCARAYLNGDGSGRGSHLSLYFVVMRGEFDSLLQWP FRQRVTLMLLDQSGKKNIMETFKPDPNSSSFKRPDGEMNIASGCPRFVAHSVLENAKNAYI KDDTLFLKVAVDLTDLEDL

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims. The claims presented are representative of the inventions disclosed herein. Other, unclaimed inventions are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33.

3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33.

4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33.

5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.

6. A composition comprising the polypeptide of claim 1 and a carrier.

7. A kit comprising, in one or more containers, the composition of claim 6.

8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.

9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

(a) introducing said polypeptide to said agent; and

(b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;

(b) contacting the cell with a composition comprising a candidate substance; and

(c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

(a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;

(b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and

(c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide,

wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.

19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33 or a biologically active fragment thereof.

20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 33.

21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.

22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 33.

23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 33.

24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 33.

25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 33, or a complement of said nucleotide sequence.

26. A vector comprising the nucleic acid molecule of claim 20.

27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.

28. A cell comprising the vector of claim 26.

29. An antibody that immunospecifically binds to the polypeptide of claim 1.

30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.

31. The antibody of claim 29, wherein the antibody is a humanized antibody.

32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

34. The method of claim 33 wherein the cell or tissue type is cancerous.

35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and

b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 33.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 33.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.

45. The method of claim 41 wherein the cell is a mammalian cell.

46. A composition comprising a purified complex of a DAPK3 protein and a DAPK3 interacting protein, wherein said DAPK3 interacting protein is selected from the group consisting of: TEM1, CG123869-01, CG129212-01, CG125927-01, CEBPD, ATF4, CG56543-01 and Prey664111.

47. A method for determining the presence or amount in a sample of a polypeptide complex comprising DAPK3 and a DAPK3- interacting protein, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the complex; and

(c) determining the presence or amount of antibody bound to said complex,

wherein the DAPK3 interacting protein is selected from the group consisting of TEM1, CG123869-01, CG129212-01, CG125927-01, CEBPD, ATF4, CG56543-01 and Prey664111; thereby determining the presence or amount of the complex in said sample.

48. A method of treating or preventing a pathology associated with a polypeptide complex comprising DAPK3 and a DAPK3- interacting protein, the method comprising administering the complex to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

49. A composition comprising a purified complex of a TRAF5 protein and a TRAF5 interacting protein, wherein said TRAF5 interacting protein is selected from the group consisting of: LTβR, AAT62352, PIAS3, CG124499-01 and TRAF5_SV.

50. A method for determining the presence or amount in a sample of a polypeptide complex comprising TRAF5 and a TRAF5 - interacting protein, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the complex; and

(c) determining the presence or amount of antibody bound to said complex,

wherein the TRAF5 interacting protein is selected from the group consisting of LTβR, AAT62352, PIAS3, CG124499-01 and TRAF5_SV; thereby determining the presence or amount of the complex in said sample.

51. A method of treating or preventing a pathology associated with a polypeptide complex comprising TRAF5 and a TRAF5 - interacting protein, the method comprising administering the complex to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 April 2003 (17.04.2003)

PCT

(10) International Publication Number
WO 03/031571 A3

(51) International Patent Classification[7]: C07K 2/00, A61K 38/00

(21) International Application Number: PCT/US02/31357

(22) International Filing Date: 2 October 2002 (02.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/327,454	5 October 2001 (05.10.2001)	US
60/327,917	9 October 2001 (09.10.2001)	US
60/328,029	9 October 2001 (09.10.2001)	US
60/328,056	9 October 2001 (09.10.2001)	US
60/328,849	12 October 2001 (12.10.2001)	US
60/329,414	15 October 2001 (15.10.2001)	US
60/330,142	17 October 2001 (17.10.2001)	US
60/341,058	22 October 2001 (22.10.2001)	US
60/343,629	24 October 2001 (24.10.2001)	US
60/349,575	29 October 2001 (29.10.2001)	US
60/346,357	1 November 2001 (01.11.2001)	US
60/391,342	25 June 2002 (25.06.2002)	US
10/262,445	1 October 2002 (01.10.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/327,454 (CIP)
Filed on	5 October 2001 (05.10.2001)
US	60/327,917 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,029 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,056 (CIP)
Filed on	9 October 2001 (09.10.2001)
US	60/328,849 (CIP)
Filed on	12 October 2001 (12.10.2001)
US	60/329,414 (CIP)
Filed on	15 October 2001 (15.10.2001)
US	60/330,142 (CIP)
Filed on	17 October 2001 (17.10.2001)
US	60/341,058 (CIP)
Filed on	22 October 2001 (22.10.2001)
US	60/343,629 (CIP)
Filed on	24 October 2001 (24.10.2001)
US	60/349,575 (CIP)
Filed on	29 October 2001 (29.10.2001)
US	60/346,357 (CIP)
Filed on	1 November 2001 (01.11.2001)
US	60/391,342 (CIP)
Filed on	25 June 2002 (25.06.2002)
US	Not furnished (CIP)
Filed on	1 October 2002 (01.10.2002)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ALSOBROOK, John, P., II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **CATTERTON, Elina** [FI/US]; 584 Boston Post Road, Madison, CT 06443 (US). **CHANT, John, S.** [CA/US]; 76 Peddlers Lane, Branford, CT 06405 (US). **CHAUDHURI, Amitabha** [IN/US]; 99 Harbor Avenue, Madison, CT 06443 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **GERLACH, Valerie, L.** [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). **GIOT, Loic** [FR/US]; 99 Country Way, Madison, CT 06443 (US). **GORMAN, Linda** [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). **GUO, Xiaojia** [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **MEZES, Peter, S.** [CA/US]; 7 Clark's Lane, Old Lyme, CT 06371 (US). **MILLET, Isabelle** [FR/US]; 74 Carrington Avenue, Milford, CT 06460 (US). **OOI, Chean, Eng** [US/US]; 14 Flax Mill Hollow, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **RIEGER, Daniel, K.** [DE/US]; 10A McKinnel Court, Branford, CT 06405 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street, Number 1, New Haven, CT 06511 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **ZERHUSEN, Bryan, D.** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **ZHONG, Haihong** [CN/US]; 2269 Long Hill Road, Guilford, CT 06437 (US). **ZHONG, Mei** [CA/US]; 45

[Continued on next page]

(54) Title: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies that immunospecifically bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the novel polypeptide, polynucleotide, or antibody specific to the polypeptide. Vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same are also included. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 03/031571 A3

Harrison Avenue, Apartment 1B, Branford, CT 06405 (US).

(74) **Agents: ELRIFI, Ivor, R.** et al.; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) **Date of publication of the international search report:**
31 July 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/31357

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C07K 2/00; A61K 38/00
US CL : 530/350; 514/2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 530/350; 514/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MEDLINE, CAPLUS, AGRICOLA, EMBASE, BIOSIS, WPIDS, EAST, GENBANK, EMBL, SWISSPROT

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Database GenBank Accession Number AL524817, February 13, 2001. AL524817 encodes a polypeptide that is 100% identical to SEQ ID NO:2.	1-3, 5, 6

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
15 May 2003 (15.05.2003)

Date of mailing of the international search report
13 JUN 2003

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
David J. Steadman
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/31357

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-8 reciting SEQ ID NO:2

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Groups I-XXXIII, claim(s) 1-8 and 36-45, drawn to an isolated polypeptide, a composition comprising a polypeptide, a kit comprising a composition, the first claimed method of making, and the first claimed method of use, i.e., the use of a therapeutic in the manufacture of a medicament, wherein the therapeutic is a polypeptide. Group I consists of SEQ ID NO:2, Group II consists of SEQ ID NO:4, Group III consists of SEQ ID NO:6, ...and Group XXXIV consists of SEQ ID NO:66.

Groups XXXIV-LXVI, claim(s) 20-28, drawn to an isoated nucleic acid, a vector comprising a nucleic acid, and a host cell comprising said vector. Group XXXIV consists of SEQ ID NO:1 or a nucleic acid encoding SEQ ID NO:2, Group XXXV consists of SEQ ID NO:3 or a nucleic acid encoding SEQ ID NO:4, Group XXXVI consists of SEQ ID NO:5 or a nucleic acid encoding SEQ ID NO:6, ...and Group LXVI consists of SEQ ID NO:65 or a nucleic acid encoding SEQ ID NO:66.

Groups LXVII-XCIX, claim(s) 9 and 29-31, drawn to an antibody that binds to a polypeptide and the first claimed method of use, i.e., a method for determining the presence or amount of a polypeptide. Group LXVII consists of an antibody that binds SEQ ID NO:2, Group LXVIII consists of an antibody that binds SEQ ID NO:4, Group LXIX consists of an antibody that binds SEQ ID NO:6, ...and Group XCIX consists of an antibody that binds SEQ ID NO:66.

Groups C-CXXXII claim(s) 10, drawn to a method for determining the presence of or predisposition to a disease associated with altered levels of a polypeptide. Group C consists of methods reciting SEQ ID NO:2, Group CI consists of methods reciting SEQ ID NO:4, Group CII consists of methods reciting SEQ ID NO:6, ...and Group CXXXII consists of methods reciting SEQ ID NO:66.

Groups CXXXIII-CLXV, claim(s) 11 and 12, drawn to a method of identifying an agent that binds to a polypeptide. Group CXXXIII consists of methods reciting SEQ ID NO:2, Group CXXXIV consists of methods reciting SEQ ID NO:4, Group CXXXV consists of methods reciting SEQ ID NO:6, ...and Group CLXV consists of methods reciting SEQ ID NO:66.

Groups CLXVI-CXCVIII, claim(s) 13, drawn to a method for identifying a potential therapeutic agent for use in treatment of a pathology. Group CLXVI consists of methods reciting SEQ ID NO:2, Group CLXVII consists of methods reciting SEQ ID NO:4, Group CLXVIII consists of methods reciting SEQ ID NO:6, ...and Group CXCVIII consists of methods reciting SEQ ID NO:66.

Groups CXCIX-CCXXXI, claim(s) 14 and 15, drawn to a method for screening for a modulator of activity or of latency or predisposition to a pathology. Group CXCIX consists of methods reciting SEQ ID NO:2, Group CXCX consists of methods reciting SEQ ID NO:4, Group CXCXI consists of methods reciting SEQ ID NO:6, ...and Group CCXXXI consists of methods reciting SEQ ID NO:66.

Groups CCXXXII-CCLXIV, claim(s) 16, drawn to a method of modulating the activity of a polypeptide. Group CCXXXII consists of methods reciting SEQ ID NO:2, Group CCXXXIII consists of methods reciting SEQ ID NO:4, Group CCXXXIV consists of methods reciting SEQ ID NO:6, ...and Group CCLXIV consists of methods reciting SEQ ID NO:66.

Groups CCLXV-CCXCVII, claim(s) 17-19, drawn to a method of treating or preventing a pathology. Group CCLXV consists of methods reciting SEQ ID NO:2, Group CCLXVI consists of methods reciting SEQ ID NO:4, Group CCLXVII consists of methods reciting SEQ ID NO:6, ...and Group CCXCVII consists of methods reciting SEQ ID NO:66.

Groups CCXCVIII-CCCXXX, claim(s) 32-34, drawn to a method for determining the presence or amount of a nucleic acid. Group CCXCVIII consists of methods reciting a nucleic acid encoding SEQ ID NO:2, Group CCXCIX consists of methods reciting a nucleic acid encoding SEQ ID NO:4, Group CCXCX consists of methods reciting a nucleic acid encoding SEQ ID NO:6, ...and Group CCCXXX consists of methods reciting a nucleic acid encoding SEQ ID NO:66.

Groups CCCXXXI-CCCLXIII, claim(s) 35, drawn to a method for determining the presence of or predisposition to a disease associated with altered levels of a nucleic acid. Group CCCXXXI consists of methods reciting SEQ ID NO:1 or a nucleic acid encoding SEQ ID NO:2, Group CCCXXXII consists of methods reciting SEQ ID NO:3 or a nucleic acid encoding SEQ ID NO:4, Group CCCXXXIII consists of methods reciting SEQ ID NO:5 or a nucleic acid encoding SEQ ID NO:6, ...and Group CCCLXIII consists of methods reciting SEQ ID NO:65 or a nucleic acid encoding SEQ ID NO:66.

INTERNATIONAL SEARCH REPORT

Groups CCCLXIV-CCCLXXI, claim(s) 46 and 48, drawn to a composition comprising a complex of a DAPK3 protein and a DAPK3-interacting protein and the first claimed method of use, i.e., a method of treating or preventing a pathology associated with a polypeptide complex comprising DAPK3 and a DAPK3-interacting protein by administering said complex. Group CCCLXIV consists of DAPK3 and TEM1, Group CCCLXV consists of DAPK3 and CG123869-01, Group CCCLXVI consists of DAPK3 and CG129212-01, Group CCCLXVII consists of DAPK3 and CG125927-01, Group CCCLXVIII consists of DAPK3 and CEBPD, Group CCCLXIX consists of DAPK3 and ATF4, Group CCCLXX consists of DAPK3 and CG56543-01, and Group CCCLXXI consists of DAPK3 and Prey664111.

Groups CCCLXXII-CCCLXXIX, claim(s) 47, drawn to a method for determining the presence or amount in a sample of a polypeptide complex comprising DAPK3 and a DAPK3-interacting protein. Group CCCLXXII consists of DAPK3 and TEM1, Group CCCLXXIII consists of DAPK3 and CG123869-01, Group CCCLXXIV consists of DAPK3 and CG129212-01, Group CCCLXXV consists of DAPK3 and CG125927-01, Group CCCLXXVI consists of DAPK3 and CEBPD, Group CCCLXXVII consists of DAPK3 and ATF4, Group CCCLXXVIII consists of DAPK and CG56543-01, and Group CCCLXXIX consists of DAPK and Prey664111.

Groups CCCLXXX-CCCLXXXIV, claim(s) 49 and 51, drawn to a composition comprising a complex of a TRAF5 protein and a TRAF5-interacting protein and the first claimed method of use, i.e., a method of treating or preventing a pathology associated with a polypeptide complex comprising TRAF5 and a TRAF5-interacting protein by administering said complex. Group CCCLXXX consists of TRAF5 and LTbetaR, Group CCCLXXXI consists of TRAF5 and AAT62352, Group CCCLXXXII consists of TRAF5 and PIAS3, Group CCCLXXXIII consists of TRAF5 and CG124499-01, Group CCCLXXXIV consists of TRAF5 and TRAF5_SV.

Groups CCCLXXXV-CCCLXXXIX, claim(s) 50, drawn to a method for determining the presence or amount in a sample of a polypeptide complex comprising TRAF5 and a TRAF5-interacting protein. Group CCCLXXXV consists of TRAF5 and LTbetaR, Group CCCLXXXVI consists of TRAF5 and AAT62352, Group CCCLXXXVII consists of TRAF5 and PIAS3, Group CCCLXXXVIII consists of TRAF5 and CG124499-01, Group CCCLXXXIX consists of TRAF5 and TRAF5_SV.

The inventions listed as Groups I-CCCLXXXIX do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Each of the corresponding polynucleotides, polypeptides, and antibodies are independent products with different uses and are structurally, biochemically, and biologically distinct. Additional or alternate methods of use are claimed for individual polynucleotides and polypeptides. 37 CFR 1.475(b) does not provide for unity of invention of more than a single product or more than one method of using a product as a combination of invention having unity of invention. Pursuant to 37 C.F.R. § 1.475 (d), the ISA considers that where multiple products and processes are claimed, the main invention shall consist of the first invention of the category first mentioned in the claims and the first recited invention of each of the other categories related thereto. Accordingly, the main invention (claims 1-8 reciting SEQ ID NO:2) comprises the first-recited product and methods of making and use thereof.

In the absence of any response from the Applicant, this Authority will establish the International Search Report based on the main invention. The claims drawn to the main invention are as follows: The claims of Group I, i.e., claims 1-8 reciting SEQ ID NO:2.